Journal of
APPLIED ECONOMETRICS
http://jae.wiley.com

5 April 2018

Dr Paolo Santucci de Magistris
Department of Economics and Business Economics
Aarhus BSS.
Fuglesangs Allé 4
DK-8210, Aarhus V

Dear Dr. Santucci de Magistris,

I am pleased to confirm that we will publish your paper, "Indirect inference with time series observed with error" in a forthcoming issue of the *Journal of Applied Econometrics.*

Congratulations on a fine piece of work!

Thank you for using the *Journal of Applied Econometrics* as an outlet for your work. I look forward to seeing it in print.

Yours sincerely

Barbara Rossi

Barbara Rossi
Editor

Proofs
Page proofs will be sent to you at the following address directly from the publishers:

eduardo.rossi@unipv.it

If this is not the correct address, please inform the Editorial Office immediately. The proofs should be received in 4-6 weeks and should be returned within a week of receipt. Authors are reminded that, at this stage, corrections should be kept to an absolute minimum. Alterations by authors may result in charges.

Illustrations
With regard to illustrations, authors should take care that original line drawings, not photocopies, are submitted in a form suitable for immediate reproduction. Drawings should be about twice the final size and the lettering must be clear and 'open' and large enough to be reproduced in the same proportion. Graphs provided on disk in colour will be published in colour in the on-line editions of the *Journal of Applied Econometrics*. A limited number of colour figures will be included in the printed Journal at the editor's discretion.

Copyright
Your article cannot be published until you have signed the appropriate license agreement. You will receive an email from Wiley's Author Services system which will ask you to log in and will present you with the appropriate license for completion.

Abstracts
Authors should provide an abstract of their paper which should cover the main aims, methods and findings of their research. This abstract must not exceed 100 words.

Offprints
Free access to the final PDF offprint of your article will be available via Author Services only. Please therefore sign up for Author Services if you would like to access your article PDF offprint and enjoy the many other benefits the service offers. Further reprints and copies of the journal may be ordered. There is no page charge to authors.

Indirect inference with time series observed with error*

Eduardo Rossi[†] Paolo Santucci de Magistris[‡]

March 23, 2018

Abstract

We propose the indirect inference estimator as a consistent method to estimate the parameters of a structural model when the observed series are contaminated by measurement error by considering the noise as a structural feature. We show that the indirect inference estimates are asymptotically biased if the error is neglected. When the condition for identification is satisfied, the structural and measurement error parameters can be consistently estimated. The issues of identification and misspecification of ME are discussed in detail. We illustrate the reliability of this procedure in the estimation of stochastic volatility models based on realized volatility measures contaminated by microstructure noise.

Keywords: Indirect inference, measurement error, misspecification, identification, stochastic volatility models
J.E.L. classification: C13, C15, C22, C58

*We are grateful to Torben Andersen, Andrea Barletta, Eric Ghysels, Roberto Renò, Viktor Todorov, two anonymous referees and the participants to the 2014 Conference on Indirect Estimation Methods in Finance and Economics (Constance, Germany) and SoFiE Conference 2015 (Aarhus, Denmark) for useful comments on previous versions of this paper. Eduardo Rossi acknowledges financial support from the MIUR PRIN project MISURA - Multivariate Statistical Models for Risk Assessment. Paolo Santucci de Magistris acknowledges financial support from CREATES - Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78).

[†]Department of Economics and Management, University of Pavia. Via San Felice 5, 27100 Pavia, Italy. E-mail: `eduardo.rossi@unipv.it`.

[‡]Department of Economics and Finance, LUISS University. Viale Romania 32, 00197 Roma, Italy. CREATES, Aarhus University, Fuglesangs Alle 4, 8210, Aarhus, Denmark. E-mail: `sdemagistris@econ.au.dk`.

1 Introduction

A common feature of many economic and financial time series is that they are recorded with errors or frictions. In macroeconomics, most series are the result of complicated processes of aggregation and are not observed exactly. For instance, the error in the measurement of GDP is a well-known problem, see Aruoba et al. (2016). In financial markets the transaction prices differ from the efficient ones when sampled at high frequency due to the features of the trading process. Regressions for time series models with errors in measurement have been discussed by Hannan (1963), Grether and Maddala (1973), Robinson (1986) and Tanaka (2002) among others. Identifiability problems for such models appear in Maravall (1979), Anderson and Deistler (1984), Nowak (1985), Solo (1986) and Chanda (1995). Staudenmayer and Buonaccorsi (2005) study the estimation of parameters in autoregressive models when measurement error (ME) is uncorrelated but possibly heteroskedastic. Hansen and Lunde (2014) study the properties of instrumental variables, such as the lagged observed series, showing that they are very weak when the signal is not persistent. Recently, more general hypotheses on ME have been considered, see Komunjer and Ng (2014) and Song et al. (2015).

In this paper, we focus on the estimation of the parameters of pure time series models when the data are observed with errors. Our methodology to deal with the errors-in-variables problem is based on indirect inference (II henceforth) and is valid also when the likelihood function (or any other criterion function that might form the basis of estimation) is analytically intractable or too difficult to evaluate. II consists of two stages. First, an auxiliary model is estimated on the observed data. Then an analytical or simulated mapping, called *binding function*, of the structural model parameters to the auxiliary statistic is calculated. II calibrates the parameters of the structural model to minimize the distance between the estimated parameters of the auxiliary models. The methodology has been introduced in the econometric literature by Smith (1993), Gouriéroux et al. (1993), Bansal et al. (1995), Gallant and Tauchen (1996), and is surveyed in Gouriéroux and Monfort (1996) and Jiang and Turnbull (2004). For example, the II methods have been successfully employed in the estimation of continuous-time models for asset prices and volatility where the transition density functions are often unknown and the stochastic volatility (SV) process is not directly observable, see among others Gallant et al. (1997), Chernov et al. (2003), and, more recently, Gagliardini et al. (2017).

We first characterize the asymptotic bias of the II estimator, which would arise if we neglect the presence of ME in the observed time series. Indeed, the noise produces an inconsistent functional estimator of the theoretical binding function and consequently an inconsistent estimator of the structural parameters. The asymptotic bias is a function of the distance between the *true* binding function and the *pseudo* binding function that results from neglecting ME parameters. In general, the econometric techniques dealing with ME do not consider the latter as a structural feature, but, rather, as a disturbance caused by the noisy observation of the economic variables, whose impact on the estimates of the model parameters must be made asymptotically negligible by means of a robust estimator. For instance, the robust estimator of SV models based on noisy squared returns studied in Hurvich et al. (2005) explicitly accounts for the impact of the noise by adding an extra parameter in the semi-parametric estimation of the long-memory parameter.

Instead, our contribution is to directly model ME by treating it as a structural feature. The solution is to consider the nuisance parameters of the noise distribution among those to be estimated. This requires to simulate trajectories from the structural model and to perturb them by ME. This represents a straightforward solution to the inconsistency of structural parameters estimates caused by the contamination of ME. The main advantage of this approach is that it is fully built within the II framework, so that consistency and asymptotic normality of the estimator are guaranteed under correct specification of ME process and identification. The issues of identification and misspecification of ME are discussed in detail. First, based on the encompassing principle, we show that II can still be consistent for the parameters of interest when the conditional distribution of ME is misspecified, as long as the structural model for the observed series encompasses the auxiliary. Second, we stress that, while the II framework provides a general setup to tackle the problem of ME, choosing an auxiliary model able to identify both the structural and the noise parameters may be non trivial and must be assessed on a case by case basis.

The proposed methodology is potentially valid in a very large number of situations, see for repeated events models Jiang et al. (1999), and it also provides a formal justification for the application of II to panel data with ME as in Guvenen and Smith (2014). The II method adopted in Pastorello et al. (2000) could be extended to account for ME associated with the fact that only the bid and ask option prices are available. Alternatively, the II estimator could be adopted when the structural model relates currently observed economic variables with the expectations formed in the previous periods, which are subject to ME by construction. For instance, if we are interested in studying how the expectations of future inflation can stimulate current consumption, we would need to account for ME and tackle the associated endogeneity problem. Thus, the robust II technique outlined in this paper could represent a valid alternative to the classic limited-information econometric methods outlined in Mavroeidis et al. (2014) for the study of the role of inflation expectations in the New Keynesian Phillips curve.

In this study, we employ II to estimate continuous-time SV models based on realized variance (RV) where the efficient price is possibly contaminated by the microstructure noise (MN). When the variance is generated by the Heston (1993) model, then the binding function relative to the HAR-RV model of Corsi (2009) can be written in terms of the SV parameters so that the identification condition can be formally verified. The analysis is further extended to the case of leverage, drift and price jumps. We show that all parameters, including those of MN, are identifiable by a multivariate auxiliary model for daily returns, RV and signed jump variations. In general, ARMA(r, l) with $r > l$ contaminated by i.i.d. noise with non-zero variance, the condition for local identification is satisfied by an autoregressive auxiliary model with $m > r + l$ lags.

Finally, we present an empirical study based on the RV series of JP Morgan to see how the technique so far discussed works in practice. The analysis corroborates the evidence emerged in the theoretical study and highlights the advantages and limits of the proposed methodology. In particular, neglecting MN makes the estimates of the SV parameters highly dependent on the choice of the sampling frequency adopted in the construction of RV. Indeed, it is not possible to reconcile the estimates of the auxiliary parameters obtained with RV based on log-returns sampled at 5 seconds and 5 minutes, unless MN is explicitly modeled. Moreover, the advantage of using the data at very high frequency is that they are informative on the generation of financial prices including the trading mechanism which is responsible for MN (e.g. bid-ask bounds and decimalization). On the contrary, sampling at lower frequencies does not provide sufficient information on the price generation mechanism and makes the identification of price jumps very difficult. This might also explain why in previous studies, based on daily returns only, the jump parameters were weakly identified, and hence were constrained, see e.g. Chernov et al. (2003). It turns out that, when the signed jump variation is adopted to disentangle price jumps from volatility dynamics and MN is in the form of a bid-ask spread, the fit of the model is dramatically improved.

The paper is organized as follows. Section 2 discusses the effect of ME on the II estimates of the structural parameters in the pure time series case and illustrates the conditions for consistency. Section 3 deals with the relevant issue of misspecification of the conditional distribution of ME, providing a condition for the consistency of the II estimator under misspecified ME. In order to illustrate the practical implication of the results in Sections 2 and 3, we characterize the bias of the II estimator of the Ornstein-Uhlenbeck process parameters and the conditions for consistency also under misspecification. In Section 5, we theoretically study the local identification conditions for the II estimator of continuous-time SV models by means of auxiliary models based on RV. Section 6 reports the results of the II estimation of SV models on JP Morgan high-frequency prices. Finally, Section 7 concludes. Appendix A contains the proofs of the propositions outlined in the paper. The Supplementary document contains additional theoretical results and examples. Moreover, it presents and discusses the results of the Monte Carlo simulations.

2 The effect of ME

Following the framework and notation of Gouriéroux et al. (1993), we first present the properties of the II method in presence of ME. The parameters of interest are those in the vector θ which characterizes the data-generating process of the unobserved series y_t. Here we consider the case of a discrete-time process y_t which is contaminated by an error term. To simplify the notation and the exposition of the results, we consider only the dependence on past values of y_t, i.e. the pure time series case.

Assumption 1 *The process* $\{y_t\}$ *is a strictly stationary and ergodic process with transition density* $f_y^0(y_t|\underline{y_{t-1}};\theta)$*, where* $\underline{y_{t-1}} = (y_{t-1},\ldots,y_{t-l})$*, that is difficult or impossible to evaluate analytically. The vector containing the true structural parameters is* $\theta_0 \in \Theta \subseteq \mathbb{R}^p$*. The parametric structural model for* y_t *is correctly specified.*

Assumption 2 *A sample of* T *observations* $\{x_t\}_{t=1}^T$ *is generated as* $x_t = g(y_t, u_t)$*,* $t = 1, 2, \ldots, T$*.*

Assumption 3 *The ME term* u_t *is supposed to be covariance stationary with a known conditional distribution, i.e.* $f_u^0(u_t|u_{t-1}, u_{t-2}, \ldots; \psi_0)$*, where* $\psi_0 \in \Psi \subseteq \mathbb{R}^h$*.*

Assumption 1 is rather standard in this framework, as II requires the process y_t to be stationary with constant moments to be asymptotically valid. The parametric structural model for y_t (conditional parametric model) is a set of conditional distributions indexed by a parameter θ, i.e.

$$\mathcal{M}_\theta(y) = \{f_y(y_t|y_{t-1},\ldots,y_{t-k};\theta), \theta \in \Theta\}.$$

The correct specification assumption means that the true conditional p.d.f. given the past of y, i.e. $f_y^0(y_t|y_{t-1},\ldots,y_{t-l};\theta)$, belongs to $\mathcal{M}_\theta(y)$ and is identifiable. Assumption 2 is quite general, it allows a non-linear mapping between the observed series x_t, the signal y_t and u_t. Generally $g(\cdot)$ is a linear/additive function or can be reduced to be linear, i.e. $x_t = y_t + u_t$. For example, suppose that the observed stock price P_t is equal to a latent efficient price times an error term with positive support, $P_t = P_t^* \cdot \tilde{\epsilon}_t$, then the efficient log-price, p_t^*, is contaminated by an additive ME term, i.e. $p_t = p_t^* + \epsilon_t$, where $\epsilon_t = \log(\tilde{\epsilon}_t)$. Assumption 3 characterizes the dynamic features of ME, which depends on a number of true nuisance parameters, contained in the vector ψ_0, and it does not exclude correlation between the signal and the noise and autocorrelation in u_t. In the following, we denote by $p_y^0 = p_y^0(\theta)$, $p_u^0 = p_u^0(\psi)$ and $p_x^0 = p_x^0(\theta,\psi)$, the true probability distributions associated with y_t, u_t and x_t.

2.1 An inconsistent estimator

We first investigate the impact of neglecting the possible presence of ME when carrying out II on θ_0 employing observations of the contaminated process x_t. The indirect inference consists of two steps: the estimation of the auxiliary (or instrumental) model and the calibration. The number of parameters in the auxiliary model, denoted by β, is q, i.e. $\beta \in \mathcal{B} \subseteq \mathbb{R}^q$, with $q \geq p$. The auxiliary model is, in general, incorrectly specified, i.e.

need not describe accurately the conditional distribution of x_t. The parameters in β can be estimated by maximizing a criterion function, $Q_T(\underline{x}_T; \beta)$, which satisfies some technical assumptions, see Gouriéroux and Monfort (1996, p.85), i.e.

$$\hat{\beta}_T = \arg\max_{\beta} Q_T(x_1, \ldots, x_T; \beta).$$

The criterion is assumed to tend asymptotically (and uniformly almost certainly) to a non-stochastic limit (see Gouriéroux et al., 1993, Assumption 2)

$$p_x^0 \lim_{T\to\infty} Q_T(x_1, \ldots, x_T; \beta) = Q_\infty(\theta_0, \psi_0, \beta).$$[1]

When the series is measured without noise, this limit depends only on the unknown auxiliary parameter β and on the true parameter of interest $\theta_0 \in \Theta \subseteq \mathbb{R}^p$. However when the series at hand is contaminated by noise this limit depends also on the true nuisance parameter vector, $\psi_0 \in \Psi \subseteq \mathbb{R}^h$, where h is the dimension of ψ_0. For example, when u_t is assumed to be an $i.i.d.N(0, \sigma_u^2)$, then $\psi_0 = \sigma_{u,0}^2$ and $h = 1$. As in Gouriéroux et al. (1993) we assume that this limit criterion is continuous in β and has a unique maximum

$$\beta_0 = \arg\max_{\beta\in\mathcal{B}} Q_\infty(\theta_0, \psi_0, \beta).$$

The binding function, i.e. the link between the auxiliary model parameters and the structural parameters, is given by

$$b(\theta, \psi) = \arg\max_{\beta\in\mathcal{B}} Q_\infty(\theta, \psi, \beta). \tag{1}$$

It follows that $\beta_0 = b(\theta_0, \psi_0)$, and $\hat{\beta}_T$ is a consistent estimator of $b(\theta_0, \psi_0)$ which is an unknown function of θ_0 and ψ_0. In the second step of the procedure, we simulate S trajectories from the DGP of y_t and the estimation of the auxiliary model is carried out on each simulated series. The auxiliary estimator based on the s-th simulated path of the signal's DGP for some θ is

$$\hat{\beta}_T^s(\theta) = \arg\max_{\beta\in\mathcal{B}} Q_T(y_1^{(s)}(\theta), \ldots, y_T^{(s)}(\theta); \beta).$$

When $T \to \infty$, $\hat{\beta}_T^s(\theta)$ converges to the solution of the limit problem

$$\tilde{b}(\theta) = \arg\max_{\beta\in\mathcal{B}} Q_\infty(\theta, \beta),$$

where $Q_\infty(\theta, \beta)$ is the limit under p_y^0 of $Q_T(y_1^{(s)}(\theta), \ldots, y_T^{(s)}(\theta); \beta)$, so that $p_y^0 \lim_{T\to\infty} \hat{\beta}_T^s(\theta) = \tilde{b}(\theta)$. Therefore, $\hat{\beta}_T^s(\cdot)$ is an inconsistent functional estimator of $b(\theta, \psi)$. The indirect in-

[1]We denote by $p_x^0 \lim_{T\to\infty}$ the limit in probability (with respect to p_x^0) when T goes to infinity.

ference estimator of θ is found by minimizing the distance between $\hat{\beta}_T$ and $\frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_T^s(\theta)$ under a metric given by the positive definite matrix Ω_T, as $\hat{\theta}_{ST} = \arg\min_\theta \Xi_T(\theta)$, where $\Xi_T(\theta) = \arg\min_\theta \left\|\hat{\beta}_T - \hat{\beta}_{ST}\right\|^2_{\Omega_T}$, where $\hat{\beta}_{ST} = \frac{1}{S}\sum_{s=1}^{S}\hat{\beta}_T^s(\theta)$. The estimator $\hat{\theta}_{ST}$ depends on the data via $\hat{\beta}_T$, and thus on the nuisance parameter ψ. Indeed, the limit of $\hat{\beta}_T$ as $T \to \infty$ is $b(\theta_0, \psi_0)$. Instead, $\hat{\beta}_{ST} \xrightarrow{p_y^0} \tilde{b}(\theta)$. This discrepancy induces an asymptotic bias in the indirect inference estimator. The following proposition establishes the inconsistency of $\hat{\theta}_{ST}$ and shows the components of the asymptotic bias, where the short-hand notation $\frac{\partial\tilde{b}(\theta_0)}{\partial\theta'}$ indicates $\left.\frac{\partial\tilde{b}(\theta)}{\partial\theta'}\right|_{\theta=\theta_0}$.

Proposition 1 *If Assumptions 1-3 hold and $\Omega_T \xrightarrow{p} \Omega$, a deterministic positive definite matrix; $\frac{\partial\hat{\beta}_{ST}}{\partial\theta'} \overset{p_x^0}{\to} \frac{\partial\tilde{b}(\theta)}{\partial\theta'}$ as $S \to \infty$ uniformly in $N_\epsilon = \{\theta; \|\theta - \theta_0\| < \epsilon\}$; then:*

$$p_x^0 \lim_{T\to\infty} \hat{\theta}_{ST}(\Omega_T) = \theta_0 + \left[\frac{\partial\tilde{b}'(\theta_0)}{\partial\theta}\Omega\frac{\partial\tilde{b}(\theta_0)}{\partial\theta'}\right]^{-1}\frac{\partial\tilde{b}'(\theta_0)}{\partial\theta}\Omega(b(\theta_0,\psi_0) - \tilde{b}(\theta_0)). \tag{2}$$

where $\frac{\partial\tilde{b}(\theta_0)}{\partial\theta'}$ is full column rank.

When $p = q$, $p_x^0 \lim \hat{\theta}_{ST} = \theta_0 + \left[\frac{\partial\tilde{b}(\theta_0)}{\partial\theta'}\right]^{-1}[b(\theta_0,\psi_0) - \tilde{b}(\theta_0)]$. This makes clear that the term responsible for the distortion and the inconsistency of $\hat{\theta}_{ST}$, i.e. $\left[\frac{\partial\tilde{b}(\theta_0)}{\partial\theta'}\right]^{-1}[b(\theta_0,\psi_0) - \tilde{b}(\theta_0)]$, does not vanish asymptotically since $b(\theta_0,\psi_0)$ depends also on ψ_0. Section 1.1 in the Supplementary document provides more details on the first and second order bias of the II estimator when ME is neglected and $p = q$. Notably, this result is general and it holds for any misspecification that involves nuisance parameters, ψ. The solution to this inconsistency is therefore based on the idea of treating ME as a potential source of misspecification, thus considering it as a structural feature, see Section 2.2.

2.2 A consistent estimator

So far, we have shown that neglecting ME generates a bias in the II estimator, as a result of an inconsistent functional estimation of the binding function. Therefore, a straightforward solution to the inconsistency caused by the presence of ME is to consider the nuisance parameters ψ among the structural parameters that need to be estimated. The $(p+h)\times 1$ parameter vector to be estimated is now denoted by $\zeta = (\theta', \psi')'$. The parameter space of ζ is $\mathcal{Z}$. The auxiliary model is characterized by a criterion function $Q_T(\underline{x}_T, \beta)$, where $\beta \in \mathcal{B}$ with $\mathcal{B}$ compact subset of $\mathbb{R}^q$, with $q \geq p + h$. The proposed II procedure requires that we simulate $y_t^{(s)}$ from the structural model and ME from its conditional density. The contaminated artificial series, i.e. $x_t^{(s)} = y_t^{(s)} + u_t^{(s)}$, are used in place of $y_t^{(s)}$, thus

$$\hat{\beta}_T^s(\zeta) = \arg\min_\beta Q_T(x_1^{(s)}(\zeta), \ldots, x_T^{(s)}(\zeta); \beta).$$

so that the knowledge of the functional form of the conditional distribution of u_t is crucial when we want to simulate from it. The estimated binding function, $\hat{b}(\zeta)$, which now explicitly depends on θ and ψ, is used to match both sets of parameters.

Similarly to Gouriéroux et al. (1993), we assume the following regularity conditions,

Assumption 4 *i. The normalized function $Q_T(\underline{x}_T^{(s)}(\zeta), \beta)$ uniformly converges in (ζ, β) to a deterministic function $Q_\infty(\zeta, \beta)$ when T diverges.*

ii. The limit function $Q_\infty(\zeta, \beta)$ has a unique maximum with respect to β. The maximum is $b(\zeta) = \arg\max_{\beta \in \mathcal{B}} Q_\infty(\zeta, \beta)$.

iii. The functions $Q_T(\underline{x}_T^{(s)}(\zeta), \beta)$ and $Q_\infty(\zeta, \beta)$ are differentiable with respect to β.

iv. The only solution of the asymptotic first order condition is associated with $\beta_0 = b(\zeta_0)$.

v. $\frac{\partial b(\zeta_0)}{\partial \zeta'}$ is a full-column rank matrix (local identification).

Assumption 4.v guarantees that ζ is locally identified, when ζ is such that $||\zeta - \zeta_0|| < \epsilon$. To have global identification $b(\zeta)$ must be a one-to-one (injective) function, see Dhaene et al. (1998), so that the equation $\beta = b(\zeta)$ admits a unique solution at the true parameter value, ζ_0. The indirect estimator of ζ, i.e. of the structural and nuisance parameters, is obtained as

$$\hat{\zeta}_{ST} = \arg\min_{\zeta} \Xi(\zeta) \tag{3}$$

with $\Xi(\zeta) = \left\| \hat{\beta}_T - \hat{\beta}_{ST}(\zeta) \right\|^2_{\Omega_T}$. Under Assumptions 1-4 the II estimator $\hat{\zeta}_{ST}$ is consistent. Moreover, under the conditions in Gouriéroux and Monfort (1996, p.86), for $T \to \infty$ and S fixed, the II estimator $\hat{\zeta}_{ST}$ is asymptotically normal, with

$$\sqrt{T}(\hat{\zeta}_{ST} - \zeta_0) \xrightarrow{d} N\Big(0, W(S, \Omega)\Big) \tag{4}$$

where $W(S, \Omega)$ is given in Gouriéroux and Monfort (1996, p.70). This follows directly from the results in Gouriéroux et al. (1993). In other words, II provides asymptotically unbiased and normal estimates in presence of ME, if the parameters governing the latter are considered among the structural ones and pseudo-data can be simulated from the contaminated structural model. If the auxiliary model is such that it locally identifies all structural parameters, i.e. Assumption 4.v is satisfied, then standard theory applies. In Section A.4.1 of Appendix A, we analyze the local identification condition for ARMA models which is instrumental to the identification condition for the Heston SV model as discussed in Section 5.

3 Misspecification of ME

In this section, we study the conditions for the consistency of the II estimator when the conditional distribution of ME is misspecified, i.e. Assumption 3 no longer holds. In general, we don't have any knowledge of the statistical properties of u_t. ME is often modeled in such a way that can accommodate for features of the data not completely captured by the structural model. As a consequence of the misspecification of the conditional density of $\{u_t, t \in \mathbb{Z}\}$, also the conditional p.d.f. of x_t can be misspecified. We denote by $p_u^* = p_u^*(\psi)$ the misspecified distribution of u_t, and consequently with $p_x^* = p_x^*(\theta, \psi)$ the misspecified distribution of x_t. We are interested in determining the conditions which ensure a consistent estimation of the parameters of interest θ_0 despite a misspecified simulator. We focus on pseudo-true values of the parameters of p_x^*, i.e. θ_0 and $\bar{\psi}$, where θ_0 is the true unknown value of the parameters of interest. The notion of pseudo-true value is an extension of Gourieroux and Monfort (1995) terminology, since the objective criterion of the auxiliary model can be different from the log-likelihood. As in Dridi and Renault (2000), we define the encompassing condition. Given the structural model for y_t, endowed with the true unknown value θ_0, and a model for u_t parametrized with $\psi \in \Psi \subset \mathbb{R}^{h^*}$, denoted by $p_u^* = p_u^*(\psi)$, then the selected structural model for x_t encompasses the auxiliary model if there exists a $\bar{\psi} \in \Psi$ such that $b(\theta_0, \psi_0) = \tilde{b}(\theta_0, \bar{\psi})$. The following proposition states the consistency of $\hat{\theta}_{ST}$ for θ_0 when the model for ME is misspecified.

Proposition 2 *Under Assumptions 1, 2, 4.i-iii,* $\Omega_T \xrightarrow{p_x^*} \Omega$, *a deterministic positive definite matrix,* $\frac{\partial \hat{\beta}_{ST}}{\partial \zeta'} \xrightarrow{p_x^*} \frac{\partial \tilde{b}(\theta, \bar{\psi})}{\partial \zeta'}$ *as* $S \to \infty$ *uniformly in* $N_\epsilon = \{\theta; \|\theta - \theta_0\| < \epsilon\}$ *where* $\frac{\partial \tilde{b}(\theta_0, \bar{\psi})}{\partial \zeta'}$ *has full rank, and assuming that there exists a* $\bar{\psi} \in \Psi$ *such that* $b(\theta_0, \psi_0) = \tilde{b}(\theta_0, \bar{\psi})$ *with* $\tilde{b}(\theta, \bar{\psi}) = \tilde{b}(\theta_0, \bar{\psi}) \Rightarrow \theta = \theta_0$ *(partial local identification), then* $\hat{\theta}_{ST} \xrightarrow{p} \theta_0$.

Following Dridi and Renault (2000) and extending the encompassing principle by Mizon and Richard (1986) to the II framework, Proposition 2 shows that even when ME is misspecified, II can produce consistent estimates of the parameters of interest θ, if the structural model encompasses the moments to match. An obvious consequence of Proposition 2 is that when the model $p_u(\psi)$ is correctly specified, i.e. it nests $p_u^0(\psi_0)$, then $\bar{\psi} = \psi_0$. Finally, from Dridi and Renault (2000, Proposition 4.1), as $T \to +\infty$

$$\sqrt{T} \begin{bmatrix} \hat{\theta}_{ST} - \theta_0 \\ \hat{\psi}_{ST} - \bar{\psi} \end{bmatrix} \xrightarrow{d} N(0, W(S, \Omega^*(S)))$$

where $W(S, \Omega^*(S)) = \frac{\tilde{\beta}(\theta_0, \bar{\psi})'}{\partial \zeta} (\Phi_0^*)^{-1} \frac{\tilde{\beta}(\theta_0, \bar{\psi})}{\partial \zeta'}$ under the optimal choice $\Omega^*(S) = \Phi_0^*(S)^{-1}$ of Ω. Section 1.2 in the Supplementary document reports the expression for Φ_0^*.

4 An illustration

In this section, we illustrate the theoretical results outlined in Sections 2 and 3 in the case of the II estimation of the Ornstein-Uhlenbeck (OU) process under ME. This choice is motivated by the possibility of obtaining results in closed-form. The OU process is the solution of the following differential equation

$$\mathrm{d}z(t) = k(\omega - z(t))\mathrm{d}t + \sigma \mathrm{d}W(t), \quad t > 0 \tag{5}$$

where $k, \omega, \sigma \geq 0$ and $W(t)$ is a standard Brownian motion on $\mathbb{R}$. The initial value of $z(0)$ is a given random variable (possibly, a constant) taken to be independent of $\{W(t)\}_{t\geq 0}$. Denote with y_t the discrete-time realizations of the continuous-time process $z(t)$, i.e. $y_t = h(z(t); \Delta)$, where $h(\cdot; \Delta)$ is a known function and Δ is the discretization step. Let assume without loss of generality that $\Delta = 1$ so that $y_t = z_t$ is the discrete realization of $z(t)$ on the unit interval. It is well known that when y_t is observed in place of $z(t)$ standard indirect inference procedures correct for the discretization error, see Gouriéroux et al. (1993) and Broze et al. (1998). The observed series, x_t, is given by $x_t = y_t + u_t$, where u_t is supposed to be i.i.d. with $\mathrm{Var}[u_t] < \infty$ and $\mathrm{Var}[u_t] = \sigma_u^2$, independent of all leads and lags of y_t. Hence x_t is the result of the interaction between the latent continuous-time signal, $z(t)$, u_t, the function $h(\cdot; \Delta)$ and the discretization step, Δ. As in Gouriéroux and Monfort (1996), the chosen auxiliary model is the following AR(1)

$$x_t = x_{t-1} + \beta_1(\beta_2 - x_{t-1}) + \beta_3 e_t, \quad e_t \sim i.i.d.N(0,1), \tag{6}$$

where the set of auxiliary parameters is $\beta = (\beta_1, \beta_2, \beta_3)'$. Let $\theta_0 = (k_0, \omega_0, \sigma_0)'$ be the vector of unknown true model parameters and $\psi_0 = \sigma_{u,0}^2$. The probability limit of $\hat{\beta}$ is

$$b(\theta_0, \psi_0) = \begin{bmatrix} 1 - \frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2k_0\sigma_{u,0}^2} \\ \omega \\ \left[\left(\frac{\sigma_0^2}{2k_0} + \sigma_{u,0}^2\right)\left[1 + \left(\frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2k_0\sigma_{u,0}^2}\right)^2\right] - \left(\frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2k_0\sigma_{u,0}^2}\right) e^{-k_0}\frac{\sigma_0^2}{k_0}\right]^{1/2} \end{bmatrix}. \tag{7}$$

The detailed derivation is in Appendix A.3.

4.1 Neglecting ME

When the noise is neglected, the function $\tilde{b}(\theta)$ is

$$\tilde{b}(\theta) = \begin{bmatrix} 1 - e^{-k} \\ \omega \\ [\frac{\sigma^2}{2k}(1 - e^{-2k})]^{1/2} \end{bmatrix}.$$

Therefore, the leading term of the asymptotic bias in (2), $c(\theta_0, \psi_0) = b(\theta_0, \psi_0) - \tilde{b}(\theta_0)$, of the II estimator $\hat{\theta}_{ST}$ is given by

$$c(\theta_0, \psi_0) = \begin{bmatrix} e^{-k_0} - \frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2k_0\sigma_{u,0}^2} \\ 0 \\ \left\{ \left(\frac{\sigma_0^2}{2k_0} + \sigma_{u,0}^2\right)\left[1 + \left(\frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2k_0\sigma_{u,0}^2}\right)^2\right] - \left(\frac{e^{-k_0}\sigma_0^2}{\sigma_0^2 + 2\kappa_0\sigma_{u,0}^2}\right) e^{-k_0}\frac{\sigma_0^2}{k_0} \right\}^{1/2} - [(\frac{\sigma_0^2}{2k_0}(1 - e^{-2k_0}))]^{1/2} \end{bmatrix},$$

which is non null when $\sigma_{u,0}^2 > 0$. Interestingly, the estimator of the long-run mean parameter ω is not affected by ME. This is a consequence of the fact that u_t is an additive i.i.d process with zero mean, so that x_t and y_t have the same long-run mean. When $\sigma_{u,0}^2 = 0$, i.e. ME is absent, the vector $c(\theta_0, \psi_0)$ is zero and $\hat{\theta}_{ST}$ is consistent.

4.2 Correct specification of ME

We now explicitly consider ME in the II estimation and consequently the variance of ME is included among the structural parameters, i.e. $\zeta = (k, \omega, \sigma, \sigma_u^2)'$. We assume correct specification of u_t, i.e. $f_u^0(u_t|u_{t-1}, u_{t-2}, \cdots; \psi)$ is known. The binding function results to be

$$b(\zeta) = \begin{bmatrix} 1 - \frac{e^{-k}\sigma^2}{\sigma^2 + 2k\sigma_u^2} \\ \omega \\ \left[\left(\frac{\sigma^2}{2k} + \sigma_u^2\right)\left[1 + \left(\frac{e^{-k}\sigma^2}{\sigma^2 + 2k\sigma_u^2}\right)^2\right] - \left(\frac{e^{-k}\sigma^2}{\sigma^2 + 2k\sigma_u^2}\right) e^{-k}\frac{\sigma^2}{k} \right]^{1/2} \\ \frac{\sigma^2}{2k} + \sigma_u^2 \end{bmatrix}, \qquad (8)$$

where the last element of $b(\zeta)$ is $\mathrm{Var}(x_t)$. The order condition is satisfied, since $q = p + 1$. The Jacobian matrix $\frac{\partial b(\zeta_0)}{\partial \zeta'}$ has full rank for any ζ_0 in Ψ. Indeed, the discretized OU process plus noise is an AR(1) and therefore satisfies the condition of Proposition 4 in Appendix A.4.1. Hence, the auxiliary model identifies all the parameters in ζ (including the variance of ME) and II provides consistent estimates of ζ.

4.3 Misspecification of ME

In the specification process of $f_u(u_t|u_{t-1}, u_{t-2}, \cdots; \psi)$ we can incur in different types of errors. First, we could misspecify f_u. For instance, suppose that the true distribution of ME is $u_t \overset{iid}{\sim} \Gamma^*(\nu, \frac{1}{\nu})$, where Γ^* denotes a Gamma distribution centered around 0, i.e. $u_t \overset{d}{=} w_t - 1$, with $w_t \overset{iid}{\sim} \Gamma(\nu, \frac{1}{\nu})$. The shape parameter is $\nu > 0$ and the scale is $\frac{1}{\nu}$, so that $\mathrm{Var}[u_t] = \frac{1}{\nu}$. Let's assume that the chosen distribution of u_t in the simulations is misspecified, e.g. $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$. If there exists a $\bar{\sigma}_u^2 \in \Psi$ such that $\bar{\sigma}_u^2 = \frac{1}{\nu}$ (encompassing condition), then Proposition 2 applies.

Alternatively, we could misspecify the dynamics of u_t. Suppose that the true ME follows an AR(1) process, i.e. $u_t = \rho u_{t-1} + v_t$ with $v_t \sim i.i.d.(0, \sigma_v^2)$, then $\sigma_u^2 = \sigma_v^2/(1-\rho^2)$,

$\psi_0 = (\rho_0, \sigma^2_{v,0})'$ with $\sigma^2_{u,0} = \sigma^2_{v,0}/(1-\rho_0^2)$. The simulator of the ME is an MA(1), i.e. $u_t = w_t + \gamma_1 w_{t-1}$ with $\sigma^2_u = (1+\gamma^2)\sigma^2_w$. The true binding function is

$$b(\theta_0,\psi_0) = \begin{bmatrix} 1 - \frac{e^{-k_0}\sigma_0^2 + 2k_0\rho_0\sigma^2_{u,0}}{\sigma_0^2 + 2k_0\sigma^2_{u,0}} \\ \omega \\ \frac{\left[\left(\frac{\sigma_0^2}{2k_0} + \sigma^2_{u,0}\right)\left[1 + \left(\frac{e^{-k_0}\sigma_0^2 + 2k_0\rho_0\sigma^2_{u,0}}{\sigma_0^2 + 2k_0\sigma^2_{u,0}}\right)^2\right] - \left(\frac{e^{-k_0}\sigma_0^2 + 2k_0\rho_0\sigma^2_{u,0}}{\sigma_0^2 + 2k_0\sigma^2_{u,0}}\right)\left(e^{-k_0}\frac{\sigma_0^2}{k_0} + \rho_0\sigma^2_{u,0}\right)\right]^{1/2}}{\frac{\sigma_0^2}{2k_0} + \sigma^2_{u,0}} \end{bmatrix},$$

while the approximated one given by the MA(1) error is

$$\tilde{b}(\theta,\psi) = \begin{bmatrix} 1 - \frac{e^{-k}\sigma^2 + 2k\gamma\sigma^2_w}{\sigma^2 + 2k\sigma^2_u} \\ \omega \\ \frac{\left[\left(\frac{\sigma^2}{2k} + \sigma^2_u\right)\left[1 + \left(\frac{e^{-k}\sigma^2 + 2k\gamma\sigma^2_w}{\sigma^2 + 2k\sigma^2_u}\right)^2\right] - \left(\frac{e^{-k}\sigma^2 + 2k\gamma\sigma^2_w}{\sigma^2 + 2k\sigma^2_u}\right)\left(e^{-k}\frac{\sigma^2}{k} + \gamma\sigma^2_w\right)\right]^{1/2}}{\frac{\sigma^2}{2k} + \sigma^2_u} \end{bmatrix}$$

The model for x_t encompasses the auxiliary since the condition $b_0(\theta_0,\psi_0) = \tilde{b}(\theta_0,\bar{\psi})$ is satisfied for values of $\bar{\gamma}$ and $\bar{\sigma}^2_w$ such that

$$\rho_0 = \frac{\bar{\gamma}}{1+\bar{\gamma}^2}, \quad \sigma^2_{v,0} = \Big(\frac{\bar{\gamma}^4 + \bar{\gamma}^2 + 1}{1+\bar{\gamma}^2}\Big)\bar{\sigma}^2_w.$$

Finally, both f_u and the dynamics can be misspecified. For instance, if $u_t = \rho u_{t-1} + v_t$ with $v_t \overset{iid}{\sim} \Gamma^*(\nu, \frac{1}{\nu})$ and the simulator of ME is an MA(1), i.e. $u_t = w_t + \gamma_1 w_{t-1}$ with $w_t \overset{iid}{\sim} N(0,\sigma^2_w)$, the encompassing condition is again satisfied since $\nu_0 = \left(\frac{\bar{\gamma}^4+\bar{\gamma}^2+1}{1+\bar{\gamma}^2}\bar{\sigma}^2_w\right)^{-1}$. Monte Carlo simulations, reported in Section 2.1 of the Supplementary document, confirm the theoretical results outlined in this section. Notably, the loss of efficiency of the II estimator when ME is included is marginal compared to the unfeasible II estimator based on the signal.

5 Indirect estimation of SV models with RV

Originally, II has been successfully applied to the estimation of continuous-time SV processes with daily data, e.g. Gallant et al. (1997), Broze et al. (1998) and Pastorello et al. (2000) among others. The recent results in the theory of RV open the door to the estimation of continuous-time SV models based on high frequency data. Under unrealistic assumptions, e.g. the absence of microstructure noise (MN), the RV is an asymptotically unbiased and efficient estimator of integrated volatility (IV). However, the presence of MN, which is generated by structural features of financial markets (like trading rules, the bid-ask spread and the discreteness of price changes), can dramatically affect the asymptotic properties of the RV estimator. Indeed, when the sampling frequency shrinks to zero, MN obscures the IV signal, see Hansen and Lunde (2006). Consequently, the choice of the sampling frequency has important consequence on the SV estimation.

The problem of the estimation of the parameters of SV models with RV under the presence of MN has been already discussed. Corradi and Distaso (2006) derive a set of sufficient conditions for the asymptotic negligibility of ME, when the moments of the unobservable IV are replaced by the moments of the RV. In a GMM setup, Todorov (2009) discusses both the possibility of *neutralizing* the noise by sampling at low frequencies and of adopting an estimator of IV robust to MN using returns at higher frequencies, see Zhang et al. (2005) and Barndorff-Nielsen et al. (2008) among others. More recently, Creel and Kristensen (2015) explicitly deal with the problem of accounting for the intradaily sampling error when realized measures are used in the estimation of continuous-time SV models. They propose a limited information method, based on the Approximate Bayesian Computation, which is an alternative to II. Finally, Kanaya and Kristensen (2016) develop a semiparamtric estimator of SV that is theoretically robust to the contamination of MN.

In what follows, we aim at estimating SV models with high-frequency data and evaluate the adequacy of jump-diffusion SV models for returns sampled at different frequencies. Consistently with the general approach outlined in this paper, this can be carried out without neutralizing the impact of MN by using noise-robust volatility measures. Instead, we can rely on the information contained in potentially distorted but efficient estimators of IV, like RV, BPV or signed jump variation as in Barndorff-Nielsen et al. (2010) and Patton and Sheppard (2015), since we implement II with simulated trajectories from the SV model contaminated by MN. In Section 5.1 we show the local identification for different specification of the Heston SV model plus MN. This result is possible since the simulation of contaminated trajectories allows us to separately deal with the component of ME due to the discretization error (i.e. the one emerging in the asymptotic distribution involving quarticity, see e.g. Barndorff-Nielsen and Shephard, 2002b) and the one generated by MN. At low frequencies MN can be neglected and the discretization error is neutralized by II when the number of simulated trajectories diverges. However, at very high frequencies, MN cannot be ignored. The need for employing very high frequency prices is dictated by the features of the structural model. Indeed, prices at these frequencies unveil characteristics that could be hardly observable at lower frequencies. For instance, they convey relevant information to disentangle the total return variation into a diffusive and a jump component. II is particularly appealing in this context since, similarly to the GMM, it provides a distance-based criterion to evaluate if the restrictions imposed by the jump-diffusion plus MN model is compatible with returns observed at very high frequency. Moreover, it forces us to think about which features of the data (moments, auxiliary statistics) we would like to match, and to what extent these features are able to identify the structural parameters.

5.1 Local Identification

5.1.1 Baseline Heston model with MN

The Heston (1993) model is a well known continuous-time stochastic process used to describe the evolution of the volatility of an underlying asset and widely used in option pricing. Assume that $\sigma^2(t)$ follows a square root process as in Heston (1993), then

$$\mathrm{d}p^*(t) = \sigma(t)\mathrm{d}W_1(t) \tag{9}$$

$$\mathrm{d}\sigma^2(t) = \kappa(\omega - \sigma^2(t))\mathrm{d}t + \varsigma\sigma(t)\mathrm{d}W_2(t) \tag{10}$$

where $\kappa > 0$ governs the speed of mean reversion, $\varsigma > 0$ is the volatility of volatility parameter, while $\omega > 0$ is the long run mean of $\sigma^2(t)$, where the latter is the instantaneous volatility and it is independent of the process $W_1(t)$. The assumption that $\mathrm{Corr}[\mathrm{d}W_1(t), \mathrm{d}W_2(t)] = 0$, i.e. absence of leverage, is relaxed in Section 5.1.2. The condition $2\kappa\omega \geq \varsigma^2$ guarantees that the volatility process is stationary and it can never reach zero.

Now, we focus on the properties of the ex-post estimates of IV, defined as $IV_t = \int_{t-1}^{t} \sigma^2(s)\mathrm{d}s$, which cumulates the instantaneous volatility over periods of unit length. A non-parametric estimator of IV_t is $RV_t(\Delta) = \sum_{i=1}^{n} r^2_{t-1+i\Delta}$, where $n = 1/\Delta$, and $r_{t-1+i\Delta}$ are the intradaily returns over the intervals $[t-1+(i-1)\Delta; t-1+i\Delta]$, for $i = 1, \ldots, n$. Hence, Δ defines the sampling frequency. When MN is present and contaminates the high-frequency returns the observed intradaily price is observed with error, i.e.

$$p_{t,i}(\Delta) = p^*_{t,i}(\Delta) + \epsilon_{t,i} \quad \text{for} \quad t = 1, ..., T \quad \text{and} \quad i = 1, ..., n \tag{11}$$

where $p^*_{t,i}(\Delta)$ is the i-th latent efficient log-price on day t. The term $\epsilon_{t,i}$ is the noise around the true price, with mean 0 and finite fourth moment and it is assumed *i.i.d.* and independent of the efficient price. Over periods of length Δ, the log-return $r_{t,i}(\Delta) \equiv r_{t-1+i\Delta}$ is given by

$$r_{t,i}(\Delta) = \left(p^*_{t,i}(\Delta) - p^*_{t,i-1}(\Delta)\right) + (\epsilon_{t,i} - \epsilon_{t,i-1}) = r^*_{t,i}(\Delta) + u_{t,i} \tag{12}$$

with $\sigma^2_u = \mathrm{Var}\left[u_{t,i}\right] < \infty$. When there is no drift in prices, the RV_t is observed with ME, that is due both to the discretization error and MN:

$$RV_t(\Delta) = IV_t + \nu_t(\Delta), \tag{13}$$

where

$$\nu_t(\Delta) \stackrel{\mathcal{L}}{=} \eta_t(\Delta) + \sum_{i=1}^{n} u^2_{t,i} + 2\sum_{i=1}^{n} \sigma_{t,i}(\Delta) z_{t,i} u_{t,i}, \tag{14}$$

and $\eta_t(\Delta) = \sum_{i=1}^{n} \eta_{t-1+i\Delta}$ is the discretization error. Meddahi (2002) proves that $\eta_t(\Delta)$ has a zero mean (under no-drift in prices) and is heteroskedastic. The correlation between

IV and $\eta_t(\Delta)$ is zero when there is no leverage effect (Barndorff-Nielsen and Shephard, 2002b and Meddahi, 2002). When both the drift and the leverage effect are assumed to be zero, Barndorff-Nielsen and Shephard (2002a) show that the discretization error for the interval $[t-1+(i-1)\Delta, t-1+i\Delta]$ can be written as

$$\eta_{t-1+i\Delta} \overset{\mathcal{L}}{=} \sigma^2_{t,i}(\Delta)\left(z^2_{t,i}-1\right) \quad \text{for } \Delta > 0, \tag{15}$$

where $z_{t,i}$ is $i.i.d.N(0,1)$ and independent of $\sigma^2_{t,i}(\Delta) = \int_{t-1+(i-1)\Delta}^{t-1+i\Delta} \sigma^2(s)\mathrm{d}s$, the integrated variance over the i-th subinterval of length Δ. Meddahi (2003) proves that when the instantaneous volatility is a square-root process, like in the Heston (1993) model, then both IV and RV have an ARMA(1,1) representation. Bollerlsev and Zhou (2002) exploit this property to estimate the Heston parameters by GMM, without explicitly modeling the MN term. For a given $\Delta > 0$, the mean and the variance of RV are equal to

$$E[RV_t(\Delta)] = E[IV_t] + E[\nu_t(\Delta)], \qquad \mathrm{Var}[RV_t(\Delta)] = \mathrm{Var}[IV_t] + \mathrm{Var}[\nu_t(\Delta)] \tag{16}$$

with $E[IV_t] = \omega$, $E[\nu_t(\Delta)] = \Delta^{-1}\sigma^2_u$, and $\mathrm{Var}[\nu_t(\Delta)] = 2\Delta^{-1}E\left[\left(\sigma^2_{t,i}(\Delta)\right)^2\right] + 4\omega\sigma^2_u + \Delta^{-1}\left(\kappa_u - \sigma^4_u\right)$, $\kappa_u = E[u_{t,i}(\Delta)^4]$, see Rossi and Santucci de Magistris (2014). A closed form expression of the term $2\Delta^{-1}E\left[\left(\sigma^2_{t,i}(\Delta)\right)^2\right]$ as a function of the structural parameters is derived in Meddahi (2002, 2003). It follows that the variance of $RV_t(\Delta)$ is $\gamma(0) = \mathrm{Var}[RV_t(\Delta)] = 2\frac{a_1^2}{\kappa^2}[\exp(-\kappa)+\kappa-1] + \mathrm{Var}[\nu_t(\Delta)]$, where $a_1 = -\varsigma\sqrt{\frac{\omega}{2\kappa}}$, and the autocovariances of $RV_t(\Delta)$ are $\gamma(j) = a_1^2\frac{[1-\exp(-\kappa)]^2\exp(-\kappa(j-1))}{\kappa^2}$ $j > 0$.

The parameter vector of the structural model is $\zeta = (\kappa, \omega, \varsigma, \sigma^2_u)'$. A candidate auxiliary model is the HAR-RV of Corsi (2009), which is

$$x_t = \phi_1 + \phi_2 x_{t-1} + \phi_3 x^w_{t-1} + \phi_4 x^m_{t-1} + e_t, \tag{17}$$

where $x_t = RV_t(\Delta)$, $x^w_t = \frac{1}{5}\sum_{t=0}^{4} x_{t-j}$ and $x^m_t = \frac{1}{22}\sum_{t=0}^{21} x_{t-j}$, and $\phi = [\phi_1, \phi_2, \phi_3, \phi_4]'$. The following proposition establishes that ζ is locally identified when the auxiliary model is the HAR-RV.

Proposition 3 *(Identification of Heston model with MN) Let the structural model be the Heston model in* (9) *and* (10)*, with the $RV(\Delta)$ and ME as in* (13) *and* (14)*, respectively with $\Delta > 0$. The auxiliary model is the HAR-RV in* (17)*, which is an AR(22) with restrictions contained in the* (23×4) *matrix, R. The binding function results to be*

$$b(\zeta) = \begin{bmatrix} [R'Q_{ZZ}R]^{-1}R'Q_{ZX} \\ Q_{XX} - [Q_{XZ}R(R'Q_{ZZ}R)^{-1}R'Q_{ZX}] \end{bmatrix} \tag{18}$$

where Q_{ZZ}, Q_{XX}, Q_{XZ} are the moment matrices of RV and are function of $\zeta = (\kappa, \omega, \varsigma, \sigma^2_u)'$. The Jacobian matrix $\frac{\partial b(\zeta)}{\partial \zeta'}$ has full column rank for any $\zeta_0 \in \mathcal{Z}$.

At a first sight, this result seems to contradict the evidence reported in Section A.4.1 about the non-identifiability of an ARMA(r, l) plus noise when $l \leq r$. On the contrary, despite RV is the sum of an ARMA(1,1) and a noise term, identification is guaranteed in this case. The reason is that, as noted by Barndorff-Nielsen and Shephard (2002b) and Meddahi (2003), the MA parameter is restricted and dependent on the autoregressive root. The consequence is that the AR and MA parameters of the ARMA(1,1) representation of RV are no more *functionally independent*, which means that the parameter space dimension is reduced by one. For what concerns the choice of the auxiliary model, the HAR-Q of Bollerslev et al. (2016) could be an alternative solution to explicitly account for the quarticity, although not strictly necessary since II asymptotically corrects the distortion induced by the discretization error. As noted by Meddahi (2003), in absence of MN, the moving average roots of the RV converge to those of IV when $\Delta \to 0$ and hence the ARMA representations of IV and RV coincide. However, when MN is present, then the mean and the variance of ME diverge as $\Delta \to 0$, thus completely obscuring the volatility signal. Taking the limits of $E[RV_t]$, $\text{Var}[RV_t]$ and $\text{Cov}[RV_t, RV_{t-j}]$ for $\Delta \to 0$, we get

$$\lim_{\Delta \to 0} E[RV_t(\Delta)] = +\infty, \quad \lim_{\Delta \to 0} \text{Var}[RV_t(\Delta)] = +\infty \quad \lim_{\Delta \to 0} \text{Cov}[RV_t, RV_{t-j}] = \text{Cov}[IV_t, IV_{t-j}].$$

Consequently, the limit in $T \to \infty$ of the binding function in (18) diverge as $\Delta \to 0$. Instead, as stated in Proposition 3, II with ME only requires that T diverges for any $\Delta > 0$. This also represents the most realistic scenario in practice. Indeed, the sampling frequency of high-frequency data has a lower bound, so that the binding function is always associated to a finite limit in empirical applications. Clearly, choosing a small Δ would convey more information about MN, in line with Zhang et al. (2005), while the volatility signal dominates as Δ increases. Due to this trade-off, it might be recommended to adopt an auxiliary model based on realized measures obtained from returns sampled at different frequencies, as discussed below.

5.1.2 Leverage

Now, we assume that in the Heston SV model (9)-(10), $\text{Corr}[\mathrm{d}W_1(t), \mathrm{d}W_2(t)] = \rho \mathrm{d}t$, namely the presence of leverage. In this case, $\text{Var}[RV_t(\Delta)] = \text{Var}[IV_t] + \text{Var}[\nu_t(\Delta)] + 2\,\text{Cov}[IV_t, \nu_t(\Delta)]$. Meddahi (2002, Proposition 4.2) shows that

$$\text{Var}[\eta_{t-1+i\Delta}] = 4\Big(\frac{\omega^2 \Delta^2}{2} + \frac{a_1^2}{\kappa^2}(\exp(-\kappa\Delta) - 1 + \kappa\Delta)\Big) + 8\rho^2 a_1 e_{1,1} \frac{e_{1,0}}{\kappa}\left[\frac{\Delta}{\kappa} + \frac{\exp(-\kappa\Delta)(\kappa\Delta)}{\kappa^2}\right]$$

where $e_{1,0} = \sqrt{2\kappa\omega}, e_{1,1} = \varsigma$. The additional term in $\text{Var}[RV_t(\Delta)]$ is the covariance between IV_t and $\nu_t(\Delta)$ which is equal to $\text{Cov}(IV_t, \nu_t(\Delta)) = 2\Delta^{-1}\rho a_1 e_{1,1} \frac{e_{1,0}}{\kappa}\left[\frac{\Delta}{\kappa} + \frac{\exp(-\kappa\Delta)(\kappa\Delta)}{\kappa^2}\right]$.

As noted by Meddahi (2002), the correlation between the noise and the integrated volatility tends to zero very quickly as one increases the frequency of intra-daily observations. Therefore, similarly to Bandi and Renò (2012, p.110), we suggest to identify the contemporaneous leverage parameter by exploiting the contemporaneous covariance between the daily returns and $RV_t(\Delta)$, that is given by

$$\mathrm{Cov}(r_t, RV_t(\Delta)) = \rho\frac{a_1 e_{1,0}}{\kappa^2}[\exp(-\kappa) - 1 + \kappa] + \frac{\rho}{\Delta}\frac{a_1 e_{1,0}}{\kappa^2}[\exp(-\kappa\Delta) - 1 + \kappa\Delta],$$

since $\mathrm{Cov}(r_t, \sum_{i=1}^n u_{t,i}^2) = 0$ as daily observed returns can be assumed unaffected by MN, without loss of generality, while for a small Δ, $[\exp(-\kappa\Delta) - 1 + \kappa\Delta] \sim \kappa^2\Delta^2/2$ so that the second term doesn't depend on ρ when $\Delta \to 0$. Since the parametric expression of $\mathrm{Cov}(r_t, RV_t(\Delta))$ doesn't involve any MN parameter, the leverage parameter ρ can be isolated from the noise and separately identified by the contemporaneous covariance between daily returns and RV, while the other SV parameters and the variance of MN can be identified by a HAR type auxiliary model as in (17), as shown in Proposition 3.

5.1.3 Drift and price jumps

We consider an extension of the Heston SV model that allows for non-zero drift and jumps in prices. We assume that the efficient price process $p^*(t)$ follows an Itô semimartingale, see Bates (1996), namely

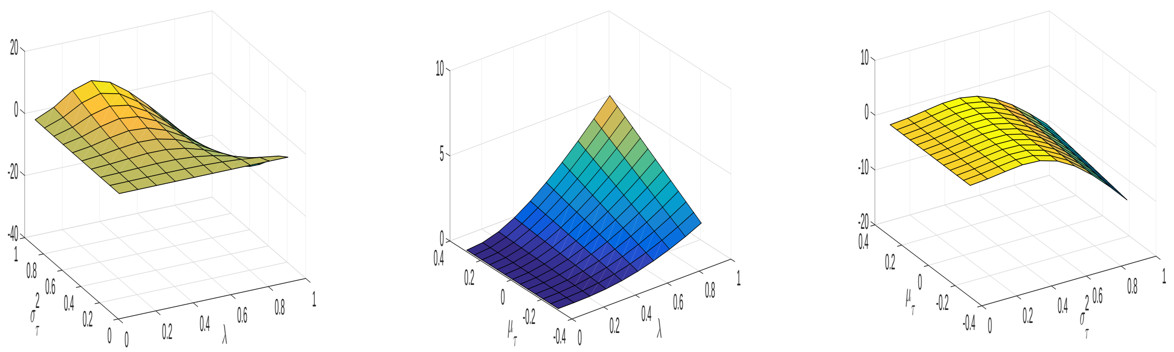
$$\mathrm{d}p^*(t) = m(t)\mathrm{d}(t) + \sigma(t)\mathrm{d}W_1(t) + \tau(t)\mathrm{d}N(t), \tag{19}$$

which implies that the efficient log-return over an interval of length Δ is $r_{t,i}^*(\Delta) = \mu_{t,i} + v_{t,i} + J_{t,i}$, where $\mu_{t,i} = \int_{t-1+(i-1)\Delta}^{t-1+i\Delta} m(s)\mathrm{d}(s)$, $v_{t,i} = \int_{t-1+(i-1)\Delta}^{t-1+i\Delta} \sigma(s)\mathrm{d}W_1(s)$ and $J_{t,i} = \sum_{j=1}^{N_{t,i}} \tau_j$. $N_{t,i} = N(t-1+i\Delta) - N(t-1+(i-1)\Delta)$ denotes the number of jumps in the i-th subinterval of day t. The jump size $\tau(t)$ is assumed time invariant and for the j-th jump arrival is distributed as $N(\mu_\tau, \sigma_\tau^2)$, while the jump arrival process, $N(t)$ is Poisson distributed, uncorrelated with $W_1(t)$ and $W_2(t)$, with λ, which is the average jump intensity on the unit interval $[t-1, t]$. It follows that $N_{t,i}$ is Poisson with intensity $\lambda\Delta$. We maintain the assumption of no leverage. If the efficient return is measured with MN, then $r_{t,i}(\Delta) = \mu_{t,i} + v_{t,i} + J_{t,i} + u_{t,i}$, so that $r_{t,i}^2(\Delta) = \int_{t-1+(i-1)\Delta}^{t-1+i\Delta} \sigma^2(s)\mathrm{d}s + \xi_{t,i}(\Delta)$, where $\xi_{t,i}(\Delta) = J_{t,i}^2 + \mu_{t,i}^2 + \left[v_{t,i}^2 - \int_{t-1+(i-1)\Delta}^{t-1+i\Delta} \sigma^2(s)\mathrm{d}s\right] + u_{t,i}^2 + 2\mu_{t,i}v_{t,i} + 2\mu_{t,i}J_{t,i} + 2\mu_{t,i}u_{t,i} + 2v_{t,i}J_{t,i} + 2v_{t,i}u_{t,i} + 2J_{t,i}u_{t,i}$. Assuming that the drift is constant, $m(t) = \mu\ \forall t$, and that $u_{t,i} \sim i.i.d.N(0, \sigma_u^2)$, the expected value and variance of the daily return are $E[r_t] = \mu + \lambda\mu_\tau$ and $\mathrm{Var}[r_t] = \omega + \sigma_u^2 + \lambda(\sigma_\tau^2 + \mu_\tau^2)$. The mean and variance of RV_t are functions of the structural parameters, as long as the variance of the realized signed jump variation of Barndorff-Nielsen et al. (2010). The latter is defined as $SJ_t(\Delta) = [RS_t^+(\Delta) - RS_t^-(\Delta)]$, where $RS_t^+(\Delta) = \sum_{i=1}^n r_{t-1+i\Delta}^2 \mathbb{1}(r_{t-1+i\Delta} > 0)$, $RS_t^-(\Delta) = \sum_{i=1}^n r_{t-1+i\Delta}^2 \mathbb{1}(r_{t-1+i\Delta} < 0)$ where $\mathbb{1}(\cdot)$ denotes the indicator function. The moments of $RV_t(\Delta)$, $SJ_t(\Delta)$ and of the squared jump component are reported in Section 1.5 of the Supplementary document.

Using these results, we can study the identification of the Heston model with jump prices and non-zero drift. We consider the following multivariate auxiliary model based on daily returns, RV_t and SJ_t

$$\begin{aligned} r_t &= \alpha + e_{r,t}, \\ RV_t(\Delta_1) &= \phi_1 + \phi_2 RV_{t-1}(\Delta_1) + \phi_3 RV_{t-2}(\Delta_1) + e_{RV,t}, \\ SJ_t(\Delta_1) &= e_{SJ_1,t}, \quad SJ_t(\Delta_2) = e_{SJ_2,t}, \end{aligned}$$

where Δ_1 and Δ_2 are two distinct sampling frequencies. The (8×1) vector of structural parameters with i.i.d. Gaussian MN is $\zeta = (\kappa, \omega, \varsigma, \sigma_\nu^2, \mu, \lambda, \mu_\tau, \sigma_\tau^2)'$. The number of auxiliary parameters is $q = 8$, i.e. we are in the exactly identified case. Proving that the rank of the Jacobian matrix of the binding function is full for any ζ_0 in Ψ is unfeasible as it would require an analysis in $\mathbb{R}^8$. Since we are interested in the identification of the jump parameters, we first compute the closed-form determinant of the Jacobian as a function of the structural parameters, adopting the MATLAB routines for symbolic calculus. We then set $\kappa_0, \omega_0, \varsigma_0, \sigma_{u,0}^2$ to the values adopted in the Monte Carlo study in Section 2.3 of the Supplementary document, with μ equal to 0.2, which corresponds to a 5% drift on annual basis. We then evaluate the determinant of the Jacobian of the binding function only for varying λ_0, $\mu_{\tau,0}$ and $\sigma_{\tau,0}^2$. See for an analogous analysis Canova and Sala (2009). We fix one of the three at a time to obtain the value of the determinant of the Jacobian for each combination of the other two parameters. The results are displayed in Figure 1 in the Supplementary document. From Panel a) it emerges that, for a given $\mu_{\tau,0} = -0.1$,



(a) $\mu_{\tau,0} = -0.1$ (b) $\sigma_{\tau,0}^2 = 0.5$ (c) $\lambda_0 = 0.3$

Figure 1: Determinant of the Jacobian matrix of the binding function of model (20) for different values of $\lambda \in (0, 1)$, $\sigma_{\tau,0}^2 \in (0, 1)$ and $\mu_\tau \in (-0.3, 0.3)$. Panel a) plots the determinant when $\mu_{\tau,0} = -0.1$, Panel b) plots the determinant when $\sigma_{\tau,0}^2 = 0.5$ and Panel c) plots the determinant when $\lambda_0 = 0.3$.

both λ_0 and $\sigma_{\tau,0}^2$ need to be different from zero to guarantee identification. Not surprisingly, as both λ_0 and $\sigma_{\tau,0}^2$ increase, the determinant moves away from zero. Identification is easier if the jump term has a greater impact on the return total variability. When instead

$\sigma^2_{\tau,0} = 0.5$ in Panel b), the determinant is zero for any $\mu_{\tau,0} \in (-0.3, 0.3)$ when $\lambda_0 = 0$. On the contrary, the determinant is larger than zero when $\mu_\tau = 0$ and it increases with λ_0. Similarly in Panel c) with $\lambda_0 = 0.3$, the determinant is zero for any $\mu_{\tau,0} \in (-0.3, 0.3)$ when $\sigma^2_{\tau,0} = 0$, while it increases (in absolute value) as $\sigma^2_{\tau,0}$ increases. In other words, variability in the jumps is required for identification purposes, that is both λ_0 and $\sigma^2_{\tau,0}$ need to be larger than zero to guarantee a non-singular Jacobian matrix, while $\mu_{\tau,0}$ can be equal to 0. Notably, the results do not qualitatively change when varying the values of the other SV parameters, $\kappa_0, \omega_0, \varsigma_0, \sigma^2_{u,0}$.

6 Empirical application

The empirical application presented in this section is meant to highlight both the relevance of the trade-off between sampling frequency and MN and the information content of high-frequency returns in the estimation of the SV parameters. This information is relevant when we are interested in specific features like jumps, which cannot be easily identified at low sampling frequencies. We estimate the parameters of the two-factor Heston model (TFSV henceforth) on the RV series of JPMorgan (JPM) from July 2, 2003 to June 29, 2007 using intradaily returns sampled at 5-seconds frequency, i.e. $n = 4680$. The choice of the sample period is motivated by the evidence of parameter instability for the TFSV model during the sub-prime crisis, between June-2007 until June 2009, as shown in Grassi and Santucci de Magistris (2015). The RV series is computed with returns sampled at two frequencies, 5-seconds and 5-minutes, RV^{5s} and RV^{5m} respectively. Table 1 reports the sample statistics of realized measures and daily returns. The long run mean of RV^{5s} is higher than that of RV^{5m}. Moreover, RV^{5m} is more noisy than RV^{5s}, meaning that the discretization error, denoted by $\eta_t(\Delta)$ in equation (14) seems to have a higher impact on the variance of ME while MN mainly impacts on the mean of RV^{5s}. The moments of the daily

	Mean	SD	SK	KU	AR(1)	AR(20)
$\tilde{r}_t$	0.0516	0.9019	0.0572	2.8371	-0.0074	0.0030
RV_t^{5s}	1.4073	0.6656	2.1826	11.454	0.6423	0.2861
RV_t^{5m}	1.0266	0.7381	2.9010	16.415	0.5204	0.0900
BPV_t^{5s}	0.6408	0.3464	2.7453	19.273	0.5558	0.0949
BPV_t^{5m}	0.9691	0.7239	3.0619	18.300	0.5148	0.0381
SJ_t^{5s}	-0.0102	0.1151	-1.2384	22.312	0.0632	0.0284
SJ_t^{5m}	0.0025	0.3473	-1.7984	31.585	0.0223	0.0481

Table 1: Sample statistics of $\tilde{r}_t = r_t/\sqrt{RV_t^{5m}}$, and realized measures of JPM.

de-volatized returns, $\tilde{r}_t = r_t/\sqrt{RV_t^{5m}}$, are rather close to those of the standard Gaussian distribution. Notably, the autocorrelation function of RV^{5s} is much higher than that of

RV^{5m}, as a consequence of the smaller impact of the discretization error on the variance of ME. The BPV computed from returns sampled at 5-seconds is clearly downward biased, as its mean and variance are much lower than those of BPV^{5m}. Conversely, the moments of BPV^{5m} are very close to those of RV^{5m}. The reason for the bias in BPV^{5s} is the *decimalization* effect, which induces discontinuities in the trajectories of returns sampled at very high-frequencies, so that the product $|r_{t,i-1}| \times |r_{t,i}|$ is often equal to 0, see the recent contribution of Bandi et al. (2017). For what concerns the signed jump variation, both SJ_t^{5s} and SJ_t^{5m} are centered around 0 and display little autocorrelation, while SJ_t^{5m} has more variability than SJ_t^{5s}.

6.0.1 A TFSV model with drift, jumps and leverage

We estimate the following TFSV Heston model, with drift, leverage and price jumps:

$$\begin{aligned}
\mathrm{d}p^*(t) &= \mu \mathrm{d}t + \sigma_1(t)\mathrm{d}W_1(t) + \sigma_2(t)\mathrm{d}W_2(t) + \tau(t)\mathrm{d}N(t) \\
\mathrm{d}\sigma_1^2(t) &= \kappa_1(\omega - \sigma_1^2(t))\mathrm{d}(t) + \varsigma_1\sigma_1(t)\mathrm{d}W_3(t) \\
\mathrm{d}\sigma_2^2(t) &= \kappa_2(\omega - \sigma_2^2(t))\mathrm{d}(t) + \varsigma_2\sigma_2(t)\mathrm{d}W_4(t),
\end{aligned} \tag{20}$$

where the parameters κ_1 and κ_2 govern the speed of mean reversion, while ς_1 and ς_2 determine the volatility of the volatility innovations. The parameter ω is the long-run mean of each volatility component and, as in Corsi and Renò (2012), it is assumed to be the same for both $\sigma_1^2(t)$ and $\sigma_2^2(t)$, in order to guarantee identification. $\{W_1(t) : t \geq 0\}$, $\{W_2(t) : t \geq 0\}$, $\{W_3(t) : t \geq 0\}$,$\{W_4(t) : t \geq 0\}$ are standard Brownian motions and $p^*(t)$ denotes the efficient log-price. The leverage effect depends on the parameters ρ_1 and ρ_2 such that $\mathrm{Corr}(\mathrm{d}W_1(t), \mathrm{d}W_3(t)) = \rho_1 \mathrm{d}t$ and $\mathrm{Corr}(\mathrm{d}W_2(t), \mathrm{d}W_4(t)) = \rho_2 \mathrm{d}t$. Similarly to Section 5.1.3, we assume that $N(t) \sim Pois(\lambda)$ and $\tau(t)$ is time invariant with $\tau_j \sim N(\mu_\tau, \sigma_\tau^2)$. The MN term is modeled either as an i.i.d Gaussian variable or as the bid-ask spread, generated as $p_{t,i} = p_{t,i}^* + \frac{\xi}{2}\mathbb{I}_{t,i}$, where ξ is the spread, and the order-driven indicator variables $\mathbb{I}_{t,i}$ are independently across t and i and identically distributed with $Pr\{\mathbb{I}_{t,i} = 1\} = Pr\{\mathbb{I}_{t,i} = -1\} = \frac{1}{2}$. This variable takes value 1 when the transaction is buyer-initiated, and -1 when it is seller-initiated.

The parameters of model (20) are collected in $\theta = (\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2, \mu, \rho_1, \rho_2, \lambda, \mu_\tau, \sigma_\tau^2)'$. Depending on the contamination scheme adopted, the structural parameters are collected in the vectors $\zeta_G = (\theta', \sigma_u^2)'$ and $\zeta_{BA} = (\theta', \xi)'$, where both ζ_G and ζ_{BA} are (12×1) vectors. Several restrictions of model (20) are considered, as well as alternative specifications for the auxiliary models. Table 2 reports the parameter estimates. The left part of Table 2 reports the parameter estimates for the baseline TFSV Heston model with no drift, no leverage and no jumps. If we neglect MN, the vector of structural parameter is $\theta = (\kappa_1, \kappa_2, \omega, \varsigma_1, \varsigma_2)'$. The auxiliary model adopted in this case is the univariate HARV model in (17) based on RV computed with returns sampled at 5 minutes (Model I) and 5 seconds (Model II).

	Baseline TFSV Heston						*Drift, Leverage and Jumps*		
	I	**II**	**III**	**III**	**III**	**III**	**IV**	**IV**	**IV**
κ_1	2.331^a	2.971^a	2.466^a	2.678^a	2.692^a	2.706^a	5.002^a	2.377^a	3.708^a
κ_2	0.056^a	0.028^b	0.023^a	0.040^a	0.039^a	0.037^a	0.017^a	0.011^a	0.015^a
ω	0.491^a	0.685^a	0.624^a	0.470^a	0.470^a	0.496^a	0.351^a	0.488^a	0.415^a
ς_1	1.769^a	1.430^a	1.312^a	1.776^a	1.780^a	1.717^a	4.076^a	2.035^a	3.064^a
ς_2	0.193^a	0.135^a	0.139^a	0.167^a	0.167^a	0.177^a	0.222^a	0.223^a	0.256^a
ρ_1	*	*	*	*	*	*	0.017	-0.262^a	-0.124^a
ρ_2	*	*	*	*	*	*	-0.124^a	-0.001	-0.016^b
μ	*	*	*	*	*	*	0.050^a	0.050^a	0.050^a
λ	*	*	*	*	*	*	0.217^a	0.062^a	0.085^a
μ_τ	*	*	*	*	*	*	-0.430^a	-0.344^a	-0.505^a
σ_τ^2	*	*	*	*	*	*	0.0198^a	0.1063^a	0.0446^a
ξ	*	*	*	0.0001^a	*	*	*	0.0005^a	*
σ_u^2	*	*	*	*	0.0001^a	0.0001^a	*	*	0.0001^a
γ	*	*	*	*	*	-0.403	*	*	*
$\hat{\Xi}$	0.000	0.000	126.9^a	10.94^c	11.06^c	11.01	63.55^a	25.20^a	51.17^a
df	0	0	6	5	5	4	3	2	2

Table 2: II estimates of the TFSV model with jumps and leverage under MN in (20). Several restrictions on the full model are considered. The asterisk indicates that the parameter is not estimated. The auxiliary model adopted for each structural specification is indicated in the first column. a, b and c stand for significance at 1%, 5% and 10% respectively. The last rows report the value of the criterion function in $\hat{\zeta}$, $\hat{\Xi}$, and the difference $df = q - p - h$, which represents the degrees of freedom of the χ^2 distribution. The auxiliary models are:

- **I**: Univariate HAR-RV (5-minutes):

$$\log(RV_t^{5m}) = \beta_{0,m} + \beta_{1,m}\log(RV_{t-1}^{5m}) + \beta_{2,m}\log(RV_{t-1,w}^{5m}) + \beta_{3,m}\log(RV_{t-1,m}^{5m}) + e_t^m$$

  with $\beta = \left(\beta_{0,m}, \beta_{1,m}, \beta_{2,m}, \beta_{3,m}, \sigma_{e,m}^2\right)'$, $q = 5$.

- **II**: Univariate HAR-RV (5-seconds):

$$\log(RV_t^{5s}) = \beta_{0,s} + \beta_{1,s}\log(RV_{t-1}^{5s}) + \beta_{2,s}\log(RV_{t-1,w}^{5s}) + \beta_{3,s}\log(RV_{t-1,m}^{5s}) + e_t^s$$

  with $\beta = \left(\beta_{0,s}, \beta_{1,s}, \beta_{2,s}, \beta_{3,s}, \sigma_{e,s}^2\right)'$, $q = 5$.

- **III**: Bivariate HAR-RV

$$\log(RV_t^{5m}) = \beta_{0,m} + \beta_{1,m}\log(RV_{t-1}^{5m}) + \beta_{2,m}\log(RV_{t-1,w}^{5m}) + \beta_{3,m}\log(RV_{t-1,m}^{5m}) + e_t^m$$
$$\log(RV_t^{5s}) = \beta_{0,s} + \beta_{1,s}\log(RV_{t-1}^{5s}) + \beta_{2,s}\log(RV_{t-1,w}^{5s}) + \beta_{3,s}\log(RV_{t-1,m}^{5s}) + e_t^s$$

  with $\beta = (\beta_{0,m}, \beta_{1,m}, \beta_{2,m}, \beta_{3,m}, \beta_{0,s}, \beta_{1,s}, \beta_{2,s}, \beta_{3,s}, \mathrm{vech}(\Sigma)')'$ with $\Sigma = \mathrm{Cov}([e_t^m, e_t^s])$, $q = 11$.

- **IV**: HAR-RV with returns equation, and SJ at different frequencies:

$$\begin{aligned} r_t &= \phi_0 + e_{r,t}, \\ RV_t^{5m} &= \beta_0 + \beta_1 RV_{t-1}^{5m} + \beta_2 RV_{t-1,w}^{5m} + \beta_3 RV_{t-1,m}^{5m} + e_{RV,t}, \\ SJ_t^{5m} &= e_{SJ_m,t}, \\ SJ_t^{5s} &= e_{SJ_s,t}, \\ \Sigma &= \mathrm{Cov}([e_{r,t}, e_{RV,t}, e_{SJ_m,t}, e_{SJ_s,t}]). \end{aligned}$$

  The vector of auxiliary parameters is $\beta = (\phi_0, \beta_0, \beta_1, \beta_2, \beta_3, \mathrm{vech}(\Sigma)')'$, $q = 15$.

In this case, there are 5 auxiliary parameters and 5 structural parameters, so that the model is exactly identified. The estimates of the parameters strongly depend on the sampling frequency selected to compute RV. Indeed, the long-run mean, ω, is approximately 30% higher for RV^{5s} than RV^{5m}, reflecting the differences observed in the sample statistics. The speed of mean reversion κ_2 is two times lower when RV^{5s} is used instead of RV^{5m} accommodating the higher persistence of RV^{5s}. We also consider a bivariate HAR-RV model as auxiliary (Model III), where the dependent variables are RV^{5m} and RV^{5s}. The caption of Table 2 contains details on all auxiliary specifications adopted in this section. If the presence of MN is neglected, the criterion function is minimized at $\Xi = 126.9$, which is statistically significant different from zero. This means that it is not possible to match the moments of RV^{5m} and RV^{5s} unless MN is explicitly incorporated in the model. Indeed, when the noise is modeled as an i.i.d. Gaussian random variable or with bid-ask spread, the adherence of the simulated processes to both RV^{5m} and RV^{5s} series improves significantly. In this case, the criterion function Ξ only leads to a marginal rejection of the model. Looking at the estimates of the parameters, they are almost identical both under i.i.d. Gaussian noise and bid-ask spread. As noted by Creel and Kristensen (2015), the two contamination schemes have an identical impact on the RV since, by a central limit theorem argument, they are approximately normally distributed. In particular, the long-run mean ω is around 0.46 and the other TFSV parameters are generally within the estimates obtained when the HAR model on RV^{5m} and RV^{5s} are estimated separately. Moreover, the estimates of σ_u^2 and ξ are almost identical, indicating that the two perturbation schemes have similar effects in the contamination of the latent efficient process. We have also considered the case of a MN following MA(1) dynamics. The point estimate of the MA coefficient, γ, is -0.403, but the coefficient is not significant, while the other Heston coefficients are qualitatively the same as those obtained under the i.i.d. MN scheme. Moreover, the criterion function $\hat{\Xi}$ is minimized in the same point as in the i.i.d. case, thus signaling the low statistical significance of the dynamics of MN for 5-seconds returns in the sample under investigation.

The right panel of Table 2 reports the estimation results when drift, leverage and jumps are included in the model, which requires extended auxiliary models for the estimation with II. In line with the theoretical results obtained in Section 5.1.3, we consider the following multivariate specification (Model IV),

$$
\begin{aligned}
r_t &= \phi_0 + e_{r,t}, \\
RV_t^{5m} &= \beta_0 + \beta_1 RV_{t-1}^{5m} + \beta_2 RV_{t-1,w}^{5m} + \beta_3 RV_{t-1,m}^{5m} + e_{RV,t}, \\
SJ_t^{5m} &= e_{SJ_m,t}, \quad SJ_t^{5s} = e_{SJ_s,t}, \quad \Sigma = \mathrm{Cov}([e_{r,t}, e_{RV,t}, e_{SJ_m,t}, e_{SJ_s,t}])
\end{aligned}
$$

so that the set of auxiliary parameters is $\beta = (\phi_0, \beta_0, \beta_1, \beta_2, \beta_3, \mathrm{vech}(\Sigma)')'$, which is a 15×1 vector. Based on this auxiliary model, the estimates of the structural parameters are generally significant, thus signaling good overall identification. For what concerns the parameters governing the jump term, we find that λ, μ_τ and σ^2_τ are highly significant in all cases, while the estimates of the other SV parameters are not much affected by the inclusion of jumps compared to the values obtained in absence of jumps. Interestingly, the estimates of the average jump size, μ_τ, is large and negative in all cases, meaning that price jumps are typically associated to bad news.When instead we look at the estimated second moment of the jumps, that is $E[J(t)^2] = \lambda\mu^2_\tau + \lambda(\sigma^2_\tau + \mu^2_\tau)$, we note that it is quite low and equal to 0.0143 and 0.0275, when MN is included in the structural model. In turns, this implies that the average jump contribution to the total return variability of the jump term is estimated between 1.47% and 3.31%.

Interestingly, when MN is modeled as a bid-ask spread, the estimated jump variability is close to zero, while the parameter ξ is very high, meaning that the variability that otherwise would be attributed to the jumps is instead due to the bouncing between bid and ask prices. Indeed, in absence of MN, the estimate of $E[J(t)^2]$ is 0.0529, which is associated to a relative jump contribution to return variability of 7.54%, a result in line with the values reported in Table 16 of Huang and Tauchen (2005). In this case, the jump component is responsible for a significant portion of the total return variation so that the estimates of the long-run mean, ω, are much smaller than those obtained without jumps. This implies that part of the gap between the unconditional mean of BPV^{5s} and RV^{5s} is attributed to the jump term, see Table 1. On the other hand, the fit of the model is not optimal in all cases, i.e. the criterion function is still too large compared to the χ^2 critical values. Notably, when MN is modeled as bid-ask spread, the criterion function is minimized at 25.20, while when it is modeled as i.i.d. Gaussian noise, the criterion function is minimized at 51.17. An explanation for this difference in the model fit is due to the inability of the contamination method based on i.i.d. Gaussian noise to provide a realistic setup for the generation of high frequency returns under the presence of jumps. The log-prices at very high frequencies are characterized by discreteness, due to the *decimalization* and rounding effects, making difficult to disentangle the volatility signal from MN and jumps. This misspecification is responsible for the fact that the criterion function is minimized far from zero, and its value is only marginally lower for the i.i.d. Gaussian noise than that obtained when noise is completely neglected, while the bid-ask spread seems to be more coherent with a realistic data generating process for the high-frequency returns. This evidence confirms the importance of correct specification of the contamination term to guarantee a good fit. Unfortunately, generating log-prices under a decimalization scheme induces discreteness in the observed log-price thus hiding the impact of changes in the SV parameters on the continuous dynamics of returns and volatility. Further, numerical problems arise in the estimation of the parameters covariance matrix. Finally, these results suggest that alternative structural models, for example the pure-jump Lévy processes of Barndorff-Nielsen and Shephard (2001) or trawl processes of Barndorff-Nielsen et al. (2014), may be better suited to model stock returns at very high-frequencies. This is left to future research.

7 Conclusions

This paper studies the inconsistency problem of the II estimator caused by measurement error in the observed series. We show that this inconsistency is originated by a mismatch between the binding function implied by the observed data and that obtained by simulation. We propose a general method to deal with this error-in-variable problem in the II framework. The solution is to jointly estimate the nuisance parameters of the error term distribution and the structural ones. Hence, the simulated series used to match the auxiliary parameters must be contaminated by the noise. Under standard assumptions, this estimator is consistent and asymptotically normal. Under misspecification of ME, we show that II can still be consistent, as long as the structural model for the observed series encompasses the auxiliary. A detailed analysis of the robustness/sensitivity of the II estimation to different misspecifications of ME is performed for the OU model. We prove the local identification of Heston SV model and extensions to leverage and jumps, when auxiliary models based on RV are adopted. Monte Carlo simulations show the viability of the proposed method in this framework and highlights the trade-off between bias reduction and efficiency as the sampling frequency changes. The empirical application illustrates the relevance of the proposed methodology when returns sampled at very high frequencies (e.g. 5 seconds) are used. Our empirical study highlights the limits of classical jump-diffusion SV models in fitting high-frequency data. This is not surprising since price variations at high-frequencies are discrete (due to the tick structure of the market) and are inflated by zeros.

References

Anderson, B. D. O. and Deistler, M. (1984). Identifiability in dynamic errors-in-variables models. *Journal of Time Series Analysis*, 5(1):1–13.

Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., and Song, D. (2016). Improving measurement: A measurement-error perspective. *Journal of Econometrics*, 191(2):384 – 397.

Bandi, F. M., Pirino, D., and Renò, R. (2017). Excess idle time. *Econometrica*, 85(6):1793–1846.

Bandi, F. M. and Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, 169(1):94 – 113.

Bansal, R., Gallant, A. R., Hussey, R., and Tauchen, G. (1995). Non-parametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics*, 66(1-2):251–287.

Barndorff-Nielsen, O. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of Royal Statistical Society. Series B*, 63:167–241.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2010). Measuring downside risk-realised semivariance. In et al., B., editor, *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press.

Barndorff-Nielsen, O. E., Lunde, A., Shephard, N., and Veraart, A. E. (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics*, 41(3):693–724.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of Royal Statistical Society. Series B*, 64:253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9(1):69–107.

Bollerlsev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(109):33–65.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*. Berlin:Springer-Verlag.

Broze, L., Scaillet, O., and Zakoian, J.-M. (1998). Quasi-indirect inference for diffusion processes. *Econometric Theory*, pages 161–186.

Canova, F. and Sala, L. (2009). Back to square one: identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449.

Chanda, K. C. (1995). Large sample analysis of autoregressive moving-average models with errors in variables. *Journal of Time Series Analysis*, 16(1):1–15.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1-2):225 – 257.

Corradi, V. and Distaso, W. (2006). Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies*, 73(3):635–667.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7:174–196.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*, 30:368–380.

Creel, M. and Kristensen, D. (2015). ABC of SV: Limited information likelihood inference in stochastic volatility jump-diffusion models. *Journal of Empirical Finance*, 31(C):85–108.

Dhaene, G., Gourieroux, C., and Scaillet, O. (1998). Instrumental models and indirect encompassing. *Econometrica*, 66(3):673–688.

Dridi, R. and Renault, E. (2000). Semi-parametric indirect inference. (EM/00/392).

Gagliardini, P., Ghysels, E., and Rubin, M. (2017). Indirects inference estimation of mixed frequency stochastic volatility state space models using MIDAS regressions and ARCH models. *Journal of Financial Econometrics*, 15(4):509–560.

Gallant, A., Hsieh, D., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, 81(1):159 – 192.

Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12:657–681.

Gourieroux, C. and Monfort, A. (1995). Testing, encompassing, and simulating dynamic econometric models. *Econometric Theory*, 11(2):195–228.

Gouriéroux, C. and Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford University Press.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:85–118.

Grassi, S. and Santucci de Magistris, P. (2015). It's all about volatility of volatility: Evidence from a two-factor stochastic volatility model. *Journal of Empirical Finance*, 30(0):62 – 78.

Grether, D. M. and Maddala, G. S. (1973). Errors in variables and serially correlated disturbances in distributed lag models. *Econometrica*, 41(2):255–262.

Guvenen, F. and Smith, A. A. (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6):2085–2129.

Hannan, E. J. (1963). Regression for time series with errors of measurement. *Biometrika*, 50(3/4):293–302.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24(2):127–161.

Hansen, P. R. and Lunde, A. (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*, 30(8):60–93.

Heston, S. (1993). A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–344.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3(4):456–499.

Hurvich, C. M., Moulines, E., and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica*, 73(4):1283–1328.

Jiang, W. and Turnbull, B. (2004). The indirect method: Inference based on intermediate statistics - a synthesis and examples. *Statistical Science*, 19:239–263.

Jiang, W., Turnbull, B. W., and Clark, L. C. (1999). Semiparametric regression models for repeated events with random effects and measurement error. *Journal of the American Statistical Association*, 94(445):111–124.

Kanaya, S. and Kristensen, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory*, 32(4):861–916.

Komunjer, I. and Ng, S. (2014). Measurement errors in dynamic models. *Econometric Theory*, 30:150–175.

Maravall, A. (1979). *Indentification in dynamic shock-error models*. Berlin: Spriger-Verlag.

Mavroeidis, S., Plagborg-Møller, M., and Stock, J. H. (2014). Empirical evidence on inflation expectations in the New Keynesian Phillips curve. *Journal of Economic Literature*, 52(1):124–188.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17:475–508.

Meddahi, N. (2003). ARMA representation of integrated and realized variances. *Econometrics Journal*, 6:335–356.

Mizon, G. E. and Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica*, pages 657–678.

Nowak, E. (1985). Global identification of the dynamic shock-error model. *Journal of Econometrics*, 27(2):211 – 219.

Pastorello, S., Renault, E., and Touzi, N. (2000). Statistical inference for random-variance option pricing. *Journal of Business and Economic Statistics*, 18(3):358–367.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 3:683–697.

Robinson, P. (1986). On the errors-in-variables problem for time series. *Journal of Multivariate Analysis*, 19(2):240–250.

Rossi, E. and Santucci de Magistris, P. (2014). Estimation of long memory in integrated variance. *Econometric Reviews*, 33:7:785–814.

Smith, J. A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:63–84.

Solo, V. (1986). Identifiability of time series models with errors in variables. *Journal of Applied Probability*, 23:63–71.

Song, S., Schennach, S. M., and White, H. (2015). Identification and estimation of non-separable models with measurement errors. *Quantitative Economics*, 6(3):749–794.

Staudenmayer, J. and Buonaccorsi, J. P. (2005). Measurement error in linear autoregressive models. *Journal of the American Statistical Association*, 100(471):pp. 841–852.

Tanaka, K. (2002). A unified approach to the measurement error problem in time series models. *Econometric Theory*, 18:278–296.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics*, 148(2):131 – 148.

Zhang, L., Mikland, P., and Ait-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100:1394–1411.

A Proofs

A.1 Proof of Proposition 1

The II estimator is defined as:

$$\hat{\theta}_{ST} = \arg\min_{\theta} \Xi_T(\theta) = \arg\min_{\theta} (\hat{\beta}_T - \hat{\beta}_{ST}(\theta))' \Omega_T (\hat{\beta}_T - \hat{\beta}_{ST}(\theta)) \tag{A.1}$$

with $\hat{\beta}_{ST}(\theta) = \frac{1}{S}\sum_{s=1}^{S} \hat{\beta}_T^s$. The first order condition for $\hat{\theta}_{ST}$ is

$$\frac{\partial \hat{\beta}'_{ST}(\hat{\theta}_{ST}(\Omega_T))}{\partial \theta} \Omega_T (\hat{\beta}_T - \hat{\beta}_{ST}(\hat{\theta}_{ST}(\Omega_T))) = 0. \tag{A.2}$$

Expanding (A.2) around the limit value θ_0,

$$\frac{\partial \hat{\beta}'_{ST}(\theta_0)}{\partial \theta} \Omega_T (\hat{\beta}_T - \hat{\beta}_{ST}(\theta_0)) - \frac{\partial \hat{\beta}'_{ST}(\theta_0)}{\partial \theta} \Omega_T \frac{\partial \hat{\beta}_{ST}(\theta_0)}{\partial \theta'} (\hat{\theta}_{ST}(\Omega_T) - \theta_0) = o_p(1).$$

Under $\Omega_T \overset{p_x^0}{\to} \Omega$,

$$\hat{\theta}_{ST}(\Omega) - \theta_0 = \Big[\frac{\partial \hat{\beta}'_{ST}(\theta_0)}{\partial \theta} \Omega \frac{\partial \hat{\beta}_{ST}(\theta_0)}{\partial \theta'}\Big]^{-1} \frac{\partial \hat{\beta}'_{ST}(\theta_0)}{\partial \theta} \Omega (\hat{\beta}_T - \hat{\beta}_{ST}(\theta_0)) + o_p(1).$$

Letting $S \to \infty$, then $\frac{\partial \hat{\beta}_{ST}}{\partial \theta'} \overset{p_x^0}{\to} \frac{\partial \tilde{b}(\theta)}{\partial \theta'}$ with $\frac{\partial \tilde{b}(\theta_0)}{\partial \theta'}$ full column rank. It follows that

$$p_x^0 \lim_{T\to\infty} \hat{\theta}_{ST}(\Omega) = \theta_0 + \Big[\frac{\partial \tilde{b}'(\theta_0)}{\partial \theta} \Omega \frac{\partial \tilde{b}(\theta_0)}{\partial \theta'}\Big]^{-1} \frac{\partial \tilde{b}'(\theta_0)}{\partial \theta} \Omega (b(\theta_0, \psi_0) - \tilde{b}(\theta_0)), \tag{A.3}$$

where $\tilde{b}(\theta_0) = p_x^0 \lim_{T\to\infty} \hat{\beta}_{ST}(\theta_0) = p_y^0 \lim_{T\to\infty} \hat{\beta}_{ST}(\theta_0)$, since $p_x^0 = p_y^0$ when $u_t = 0$, as in the simulated trajectories.

A.2 Proof of Proposition 2

We follow the proof of Proposition 1. Let $\zeta = (\theta', \psi')'$, and expand (A.2) around the point $\bar{\zeta}_0 = (\theta_0, \bar{\psi})$

$$\frac{\partial \hat{\beta}'_{ST}(\theta_0, \bar{\psi})}{\partial \zeta} \Omega_T (\hat{\beta}_T - \hat{\beta}_{ST}(\theta_0, \bar{\psi})) - \frac{\partial \hat{\beta}'_{ST}(\theta_0, \bar{\psi})}{\partial \zeta} \Omega_T \frac{\partial \hat{\beta}_{ST}(\theta_0, \bar{\psi})}{\partial \zeta'} ((\hat{\theta}'_{ST}, \hat{\psi}'_{ST})' - (\theta'_0, \bar{\psi}')').$$

Since $\Omega_T \overset{p_x^*}{\to} \Omega$ and $\frac{\partial \hat{\beta}_{ST}}{\partial \zeta'} \overset{p_x^*}{\to} \frac{\partial \tilde{b}(\theta, \bar{\psi})}{\partial \zeta'}$ as $S \to \infty$,

$$p_x^* \lim_{T \to \infty} \begin{bmatrix} \hat{\theta}_{ST}(\Omega_T) \\ \hat{\psi}_{ST}(\Omega_T) \end{bmatrix} = (\theta'_0, \bar{\psi}')' + \Big[\frac{\partial \tilde{b}'(\theta_0, \bar{\psi})}{\partial \zeta} \Omega \frac{\partial \tilde{b}(\theta_0, \bar{\psi})}{\partial \zeta'} \Big]^{-1} \frac{\partial \tilde{b}'(\theta_0, \bar{\psi})}{\partial \zeta} \Omega (b(\theta_0, \psi_0) - \tilde{b}(\theta_0, \bar{\psi})).$$

Premultiplying both sides by $[I_p, 0_{h \times h}]$ we get

$$p_x^* \lim_{T \to \infty} \hat{\theta}_{ST}(\Omega_T) = \theta_0 + \Lambda \frac{\partial \tilde{b}'(\theta_0, \bar{\psi})}{\partial \zeta} \Omega (b(\theta_0, \psi_0) - \tilde{b}(\theta_0, \bar{\psi})). \qquad \text{(A.4)}$$

where $\Lambda = [I_p, 0_{h \times h}] \Big[\frac{\partial \tilde{b}'(\theta_0, \bar{\psi})}{\partial \zeta} \Omega \frac{\partial \tilde{b}(\theta_0, \bar{\psi})}{\partial \zeta'} \Big]^{-1}$. Provided that $\frac{\partial \tilde{b}(\theta_0, \bar{\psi})}{\partial \zeta'}$ has full-column rank, the encompassing condition implies that there exists a $\bar{\psi} \in \Psi$ such that $b(\theta_0, \psi_0) = \tilde{b}(\theta_0, \bar{\psi})$. The consistency of $\hat{\theta}_{ST}(\Omega_T)$ follows.

A.3 Binding function of the OU model

The auxiliary model can be written as

$$x_t = \gamma_1 + \gamma_2 x_{t-1} + \gamma_3 e_t, \quad e_t \sim i.i.d. N(0, 1) \qquad \text{(A.5)}$$

with $\beta_1 = 1 - \gamma_2$, $\beta_2 = \gamma_1/(1 - \gamma_2)$ and $\gamma_3 = \beta_3$. To obtain the binding function we start with the exact discretized process of $z(t)$, which is

$$y_t = \omega(1 - e^{-k}) + e^{-k} y_{t-1} + \sigma \left(\frac{1 - e^{-2k}}{2k} \right)^{1/2} \epsilon_t \quad \epsilon_t \sim N(0, 1),$$

with $E[y_t] = \omega$, $\text{Var}[y_t] = \frac{\sigma^2}{2k}$ and $E[y_t y_{t-1}] = \omega^2 + \frac{\sigma^2 e^{-k}}{2k}$. The maximum likelihood estimators of γ_1, γ_2 and γ_3 based on x_t, are $\hat{\gamma}_1 = \frac{1}{T-1} \sum_t x_t - \hat{\gamma}_2 \frac{1}{T-1} \sum_t x_{t-1}$, $\hat{\gamma}_2 = \frac{\sum_t x_t x_{t-1} - \sum_t x_t \sum_t x_{t-1}}{\sum_t x_{t-1}^2 - (\sum_t x_{t-1})^2}$, and $\hat{\gamma}_3 = \frac{1}{T-1} \sum_t (x_t - \hat{\gamma}_1 - \hat{\gamma}_2 x_{t-1})^2$. The probability limit of $\hat{\beta}_T = (\hat{\beta}_{1,T}, \hat{\beta}_{2,T}, \hat{\beta}_{3,T})'$ is readily obtained from $\hat{\gamma}$. Consider $p \lim \hat{\gamma}_2 = p \lim \frac{\sum_t x_t x_{t-1} - \sum_t x_t \sum_t x_{t-1}}{\sum_t x_{t-1}^2 - (\sum_t x_{t-1})^2}$. The limit in probability of the terms in the numerator is $p \lim \frac{1}{T} \sum_t x_t x_{t-1} = p \lim \frac{1}{T} \sum_t y_t y_{t-1} = \omega^2 + e^{-k} \frac{\sigma^2}{2k}$ and $p \lim \frac{1}{T} \sum_t x_t \frac{1}{T} \sum_t x_{t-1} = \omega^2$, while the limit of the denominator is $p \lim \frac{1}{T} \Big\{ \sum_t (y_{t-1} + u_{t-1})^2 - \Big[\sum_t (y_{t-1} + u_{t-1}) \Big]^2 \Big\} = \text{Var}[y_t + u_t] = \text{Var}[y_t] + \sigma_u^2$. Combining the two we obtain $p \lim \hat{\beta}_1 = 1 - p \lim \hat{\gamma}_2 = 1 - \frac{e^{-k} \sigma^2}{\sigma^2 + 2k\sigma_u^2}$. Now, for $\hat{\beta}_2 = \frac{\hat{\gamma}_1}{\hat{\beta}_1}$, we have $p \lim \hat{\gamma}_1 = \omega \beta_1$, so that $p\lim \hat{\beta}_2 = \frac{\hat{\gamma}_1}{\hat{\beta}_1} = \omega$. Noting that $E[x_t] = \omega$, $E[x_t^2] = \text{Var}[y_t] + \omega^2 + \sigma_u^2$ and $E[x_t x_{t-1}] = e^{-k}\,\text{Var}(y_t) + \omega^2$ and keeping in mind that $p\lim \hat{\beta}_1 = 1 - \frac{e^{-k}\sigma^2}{\sigma^2 + 2k\sigma_u^2}$, we get the plim of $\hat{\beta}_3$, that is $p\lim \hat{\beta}_3 = \left(\frac{\sigma^2}{2k} + \sigma_u^2\right)(1 + (1-\beta_1)^2) - (1-\beta_1)e^{-k}\left(\frac{\sigma^2}{k}\right)$.

A.4 Proof of Proposition 3

Before discussing the proof of Proposition 3, we present the local identification condition for an ARMA plus noise model which is closely related to the local identification of the Heston SV model with RV.

A.4.1 Local identification of ARMA models

Chanda (1995) proposes an estimator for the ARMA plus ME model, and discusses the identification conditions. The estimator relies on preliminary estimates of the autoregressive coefficients, which make the implementation of the recursive and nonlinear conditions of the estimator computationally involved. The implementation of the II estimator is instead rather straightforward, once the condition for local identification are satisfied. Therefore, this section discusses the identification condition for a stationary ARMA(r, l) signal observed with additive ME:

$$\begin{aligned} x_t &= y_t + u_t \quad u_t \sim i.i.d.N(0, \sigma_u^2) \\ \alpha(L)y_t &= c + \varphi(L)\varepsilon_t \quad \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2). \end{aligned} \tag{A.6}$$

We assume that the roots of $\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_r L^r$ and $\varphi(L) = 1 + \varphi_1 L + \ldots + \varphi_l L^l$ all lie outside the unit circle, and there are no common roots. Note that

$$\alpha(L)x_t = c + \varphi(L)\varepsilon_t + \alpha(L)u_t \tag{A.7}$$

is an ARMA$(r$,max$\{r, l\})$. The parameter vector is $\zeta = (c, \alpha_1, \ldots, \alpha_r, \varphi_1, \ldots, \varphi_l, \sigma_\varepsilon^2, \sigma_u^2)'$. The auxiliary model is an AR(m), i.e.

$$\phi(L)x_t = \phi_0 + e_t, \tag{A.8}$$

where $E(e_t) = 0$, $E(e_t^2) = \sigma_e^2$ and $\phi(L) = (1 - \phi_1 L - \ldots - \phi_m L^m)$. The $(q \times 1)$ vector of auxiliary parameters is $\beta = (\phi_0, \phi_1, \ldots, \phi_m, \sigma_e^2)'$, with $q = (m+2) \geq p + 1 = 3 + r + l$.

Lemma 2.5 in Chanda (1995) makes clear that the identifiability of $\varphi_1, \ldots, \varphi_l$, σ_ε^2 and σ_u^2 is possible if and only if $r > l$. In the following proposition we explicit the identification condition of ζ_0 (Assumption 4.v) when the auxiliary model is an AR(m).

Proposition 4 *Let the structural model be the stationary ARMA(r,l) in* (A.6) *with ME* $u_t \sim WN(0, \sigma_u^2)$ *and the auxiliary model be the AR(m) in* (A.8) *with* $m > r + l$. *The* *binding function is*

$$b(\zeta) = \begin{bmatrix} Q_{ZZ}^{-1}Q_{ZX} \\ Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX} \end{bmatrix}$$

where Q_{ZZ}*,* Q_{XX} *and* Q_{ZX} *contain the mean, the variance and the autocovariances of the process for* x_t *in* (A.7) *up to lag* m*. Then the Jacobian matrix*

$$\frac{\partial b(\zeta)}{\partial \zeta'} = \begin{bmatrix} -(Q'_{ZX}Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1})\frac{\partial vec Q_{ZZ}}{\partial \zeta'} + Q_{ZZ}^{-1}\frac{\partial Q_{ZX}}{\partial \zeta'} \\ \frac{\partial Q_{XX}}{\partial \zeta'} - vec(Q_{ZX}Q_{XZ})'(Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1})\frac{\partial vec Q_{ZZ}}{\partial \zeta'} - 2Q_{XZ}Q_{ZZ}^{-1}\frac{\partial Q_{ZX}}{\partial \zeta'} \end{bmatrix}$$

has full column rank in $\zeta_0 \in \mathcal{Z}$ *if* $r > l$*.*

Proof 1 *Let* $q = m + 2$ *and* $p + 1 = 3 + r + l$*, with* $q \geq p + 1$*. Given the AR(m) model in* (A.8)*, the OLS estimates of* $\beta = (\phi_0, \phi_1, \ldots, \phi_m)'$ *converges in probability, when* T *diverges, to* $Q_{ZZ}^{-1}Q_{ZX}$ *(see Proposition 8.10.1 and Theorem 8.1.1 in Brockwell and Davis, 1991), while*

$$p\lim_{T\to\infty} \hat{\sigma}_e^2 = Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}$$

where

$$Q_{ZZ} = \begin{bmatrix} 1 & \mu_x & \mu_x & \ldots & \mu_x \\ \mu_x & \gamma_x(0) + \mu_x^2 & \gamma_x(1) + \mu_x^2 & \ldots & \gamma_x(m-1) + \mu_x^2 \\ \mu_x & \gamma_x(1) + \mu_x^2 & \gamma_x(0) + \mu_x^2 & \ldots & \gamma_x(m-2) + \mu_x^2 \\ \vdots & & & & \vdots \\ \mu_x & \gamma_x(m-1) + \mu_x^2 & \gamma_x(m-2) + \mu_x^2 & \ldots & \gamma_x(0) + \mu_x^2 \end{bmatrix} \quad \text{(A.9)}$$

and $Q_{XZ} = [\mu_x, \gamma_x(1) + \mu_x^2, \ldots, \gamma_x(m) + \mu_x^2] = Q'_{ZX}$*. Since we assume that* $E(u_t) = 0$*, then* $\mu_x = \mu \equiv E[y_t]$*. The variance and the autocovariances of* x_t *are* $Q_{XX} \equiv \gamma_x(0) = \gamma(0) + \sigma_u^2$ *and* $\gamma_x(k) = \gamma(k)$ *when* $k \neq 0$*, where* $\gamma(k) = \mathrm{Cov}[y_t, y_{t-k}]$*. Thus the binding function is*

$$b(\zeta) = \begin{bmatrix} Q_{ZZ}^{-1}Q_{ZX} \\ Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX} \end{bmatrix}. \quad \text{(A.10)}$$

In order to find the Jacobian matrix of $b(\zeta)$ *consider the differential for each component of* $b(\zeta)$*. Since* $\mathrm{d}(Q_{ZZ}^{-1}Q_{ZX}) = -(Q_{ZZ}^{-1}\mathrm{d}Q_{ZZ}Q_{ZZ}^{-1})Q_{ZX} + Q_{ZZ}^{-1}\mathrm{d}Q_{ZX}$*, and given that* $vec(ABCD) = (D'C' \otimes A)vec(B)$ *for suitably dimensioned matrices, where the* vec *operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other,*

$$\mathrm{d}vec(Q_{ZZ}^{-1}Q_{ZX}) = \Big[-(Q'_{ZX}Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1})\frac{\partial vec(Q_{ZZ})}{\partial \zeta'} + Q_{ZZ}^{-1}\frac{\partial Q_{ZX}}{\partial \zeta'}\Big]\mathrm{d}\zeta.$$

The differential of the second component of $b(\zeta)$ *is*

$$\mathrm{d}[Q_{XX}] - \mathrm{d}[Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}] = \mathrm{d}[Q_{XX}] - \left\{Q_{XZ}\mathrm{d}[Q_{ZZ}^{-1}]Q_{ZX}\right\} - 2\left\{Q_{XZ}Q_{ZZ}^{-1}\mathrm{d}[Q_{ZX}]\right\}$$

where $\mathrm{d}[Q_{XX}] = \frac{\partial Q_{XX}}{\partial \zeta'}\mathrm{d}\zeta$, $Q_{XZ}\mathrm{d}[Q_{ZZ}^{-1}]Q_{ZX} = -vec(Q_{ZX}Q_{XZ})'(Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1})\frac{\partial vec Q_{ZZ}}{\partial \zeta'}\mathrm{d}\zeta$, *and*

$$\left\{Q_{XZ}Q_{ZZ}^{-1}\mathrm{d}[Q_{ZX}]\right\} = Q_{XZ}Q_{ZZ}^{-1}\frac{\partial (Q_{ZX})}{\partial \zeta'}\mathrm{d}\zeta.$$

Finally, the Jacobian matrix can be written as

$$\frac{\partial b(\zeta)}{\partial \zeta'} = \begin{bmatrix} 0 \\ \frac{\partial Q_{XX}}{\partial \zeta'} \end{bmatrix} + \begin{bmatrix} -(Q_{ZX}'Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1}) & Q_{ZZ}^{-1} \\ vec(Q_{ZX}Q_{XZ})'(Q_{ZZ}^{-1} \otimes Q_{ZZ}^{-1}) & -2Q_{XZ}Q_{ZZ}^{-1} \end{bmatrix} \begin{bmatrix} \frac{\partial vec Q_{ZZ}}{\partial \zeta'} \\ \frac{\partial Q_{ZX}}{\partial \zeta'} \end{bmatrix} = A + BC$$

The rank of A *is* $p+1$ *no matter what is the value of* l *and* r*. The rank of* B *is* $m+2$ *which is larger by assumption than* $p+1$*. It follows that the rank of* $\partial b(\zeta_0)/\partial \zeta'$ *depends on the column rank of the* $((m+1)^2 + (m+1)) \times (p+1)$ *matrix* C*. The rows of the matrix* C *contain the partial derivatives of* $\mu^2 + \gamma(k), k = 0, 1, \ldots, m$ *with respect to* c*,* $\alpha = (\alpha_1, \ldots, \alpha_l)'$*,* $\varphi = (\varphi_1, \ldots, \varphi_r)'$*,* σ_ε^2 *and* σ_u^2*, i.e.*

$$\begin{bmatrix} \frac{\partial(\mu^2+\gamma_x(k))}{\partial c} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \alpha'} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \varphi'} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \sigma_\varepsilon^2} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \sigma_u^2} \end{bmatrix}$$

where

$$\frac{\partial(\mu_x^2 + \gamma_x(k))}{\partial c} = \frac{2c}{(1-\sum_{i=1}^r \alpha_i)^2}, \qquad \frac{\partial(\mu_x^2 + \gamma_x(k))}{\partial \alpha'} = 2\mu\frac{\partial \mu}{\partial \alpha'} + \frac{\partial \gamma(k)}{\partial \alpha'},$$

$$\frac{\partial(\mu_x^2 + \gamma_x(k))}{\partial \varphi'} = \frac{\partial \gamma(k)}{\partial \varphi'}, \quad \frac{\partial(\mu_x^2 + \gamma_x(k))}{\partial \sigma_\varepsilon^2} = \frac{\partial \gamma(k)}{\partial \sigma_\varepsilon^2}, \quad \frac{\partial(\mu_x^2 + \gamma_x(k))}{\partial \sigma_u^2} = \frac{\partial \gamma_x(k)}{\partial \sigma_u^2} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}.$$

First, consider the case of ARMA(1,0). The rows of C *have the following expression*

$$c_k = \begin{bmatrix} \frac{\partial(\mu^2+\gamma_x(k))}{\partial c} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \alpha_1} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \sigma_\varepsilon^2} & \frac{\partial(\mu^2+\gamma_x(k))}{\partial \sigma_u^2} \end{bmatrix}'.$$

If C *has reduced rank then* $Cw = 0$ *for* $w \neq 0$*, i.e. there exists a vector* $w = [w_1, w_2, w_3 w_4]'$ *such that* $c_k' w = 0$ *for all rows of* C*. The partial derivatives in* c_k *are*

$$\frac{\partial(\mu^2 + \gamma_x(k))}{\partial c} = \frac{2c}{(1-\alpha_1)^2}, \qquad \frac{\partial(\mu^2 + \gamma_x(k))}{\partial \alpha_1} = 2\mu\frac{\partial \mu}{\partial \alpha_1} + \left(k\alpha_1^{k-1}\gamma(0) + \alpha_1^k \frac{\partial \gamma(0)}{\partial \alpha_1}\right),$$

$$\frac{\partial(\mu^2 + \gamma_x(k))}{\partial \sigma_\varepsilon^2} = \alpha_1^k \frac{\partial \gamma(0)}{\partial \sigma_\varepsilon^2}.$$

For $k = 0$, the reduced rank condition implies

$$w_1\Big(\frac{\partial(\mu^2+\gamma_x(0))}{\partial c}\Big) + w_2\Big(\frac{\partial(\mu^2+\gamma_x(0))}{\partial \alpha_1}\Big) + w_3\Big(\frac{\partial(\mu^2+\gamma_x(0))}{\partial \sigma^2_\varepsilon}\Big) + w_4 = 0$$

while for $k > 0$

$$w_1\Big(\frac{\partial(\mu^2+\gamma_x(k))}{\partial c}\Big) + w_2\Big(\frac{\partial(\mu^2+\gamma_x(k))}{\partial \alpha_1}\Big) + w_3\Big(\frac{\partial(\mu^2+\gamma_x(k))}{\partial \sigma^2_\varepsilon}\Big) = 0.$$

Equating the two expressions above

$$w_2\Big[k\alpha_1^{k-1}\gamma(0) + \alpha_1^k\Big(\frac{\partial\gamma(0)}{\partial\alpha_1} - \frac{\partial\gamma_x(0)}{\partial\alpha_1}\Big)\Big] + w_3\frac{\partial\gamma_x(0)}{\partial\sigma^2_\varepsilon}(a^k-1) = w_4.$$

It is easy to see that w_4 varies with k which implies that cannot exist any vector $w \neq 0$ which lies in the null column subspace of C, hence in the ARMA(1,0) case the column rank of C is full. On the contrary, when $r = l = 1$, the rank of C is smaller than $p+1$. Indeed, we can show that there exists a vector $w \neq 0$ such that $Cw = 0$. For ARMA(1,1) the rows of $\partial b(\zeta)/\partial\zeta'$ consist of $\Big[\frac{\partial(\mu^2+\gamma_x(k))}{\partial c} \quad \frac{\partial(\mu^2+\gamma_x(k))}{\partial\alpha_1} \quad \frac{\partial(\mu^2+\gamma_x(k))}{\partial\varphi_1} \quad \frac{\partial(\mu^2+\gamma_x(k))}{\partial\sigma^2_\varepsilon} \quad \frac{\partial(\mu^2+\gamma_x(k))}{\partial\sigma^2_u}\Big]$, where the variance and the autocovariances are

$$\gamma_x(0) = \sigma^2_\varepsilon\frac{1+\varphi_1^2+2\alpha_1\varphi_1}{1-\alpha_1^2} + \sigma^2_u, \quad \gamma(k) = \sigma^2_\varepsilon\alpha_1^{k-1}\frac{(\alpha_1+\varphi_1)(1+\alpha_1\varphi_1)}{1-\alpha_1^2} \qquad k > 1$$

with $\frac{\partial\gamma(k)}{\partial\varphi_1} = \sigma^2_\varepsilon\alpha_1^{k-1}\frac{(1+\alpha_1\varphi_1)+\alpha_1(\alpha_1+\varphi_1)}{1-\alpha_1^2}$, $\frac{\partial\gamma(k)}{\partial\sigma^2_\varepsilon} = \alpha_1^{k-1}\frac{(\alpha_1+\varphi_1)(1+\alpha_1\varphi_1)}{1-\alpha_1^2}$. The vector w that is orthogonal to the rows of C is

$$w = (0, 0, 1, -\sigma^2_\varepsilon\frac{(1+\alpha_1\varphi_1)+\alpha_1(\alpha_1+\varphi_1)}{(\alpha_1+\varphi_1)(1+\alpha_1\varphi_1)}, w_5)',$$
$$w_5 = \sigma^2_\varepsilon\frac{(1+\alpha_1\varphi_1)+\alpha_1(\alpha_1+\varphi_1)}{(\alpha_1+\varphi_1)(1+\alpha_1\varphi_1)}\frac{\partial\gamma(0)}{\partial\sigma^2_\varepsilon} - \frac{\partial\gamma(0)}{\partial\varphi_1}.$$

For the case $r < l$ an analogous argument shows that the column rank of C is reduced. Thus when $r \leq l$ the rank of C is smaller than $p+1$.

A.4.2 Local Identification of Heston SV model

The HAR-RV model can be written as an AR(22) with linear restrictions on the autoregressive parameters

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_{22}x_{t-22} + e_t \tag{A.11}$$

where $\alpha_0 = \phi_1$, $\alpha_1 = (\phi_2 + \phi_3/5 + \phi_4/22)$, $\{\alpha_2, ..., \alpha_5\} = (\phi_3/5 + \phi_4/22)$ and $\{\alpha_6, ..., \alpha_{22}\} = (\phi_4/22)$. Let R be the 23×4 matrix with the linear restrictions, then a compact expression for α is $\alpha = R\phi$. The restricted AR(22) model in (A.11) can be estimated by OLS imposing the restriction contained in the matrix R. The OLS estimate of ϕ is

$$\hat{\phi}_T = \left[R'\Big(\sum_t z_t z_t'\Big)R\right]^{-1} R' \sum_t (z_t x_t),$$

where $z_t = (1, x_{t-1}, \ldots, x_{t-22})'$. Under standard assumptions, it can be shown that $\frac{1}{T}\sum_t (z_t z_t') \xrightarrow{p} E[z_t z_t'] \equiv Q_{ZZ}$, where Q_{ZZ} is given in (A.9) with $m = 22$, and $\mu = E[x_t]$. Therefore, $p\lim_{T\to\infty} \hat{\alpha}_T = Q_{ZZ}^{-1} Q_{ZX}$, $Q_{ZX} = E[z_t x_t] = [\mu, \mu^2 + \gamma(1), \ldots, \mu^2 + \gamma(22)]'$ and $Q_{XX} = E[x_t^2] = \gamma(0) + \mu^2$. The limit of $\hat{\phi}_T$

$$p\lim_{T\to\infty} \hat{\phi}_T = \left[R' p\lim_{T\to\infty}\Big(\sum_t z_t z_t'\Big)R\right]^{-1} R' p\lim_{T\to\infty}\sum_t (z_t x_t) = \left[R' Q_{ZZ} R\right]^{-1} R' Q_{ZX},$$

The matrix Q_{ZZ} and the vector Q_{ZX} both depend on the structural parameters ζ. The estimator of the variance of e_t is

$$\hat{\sigma}^2_{e,T} = \frac{\sum_t \hat{e}_t^2}{T}, \quad p\lim_{T\to\infty} \hat{\sigma}^2_{e,T} = Q_{XX} - Q_{XZ} R (R' Q_{ZZ} R)^{-1} R' Q_{ZX}.$$

Then the binding function results to be $b(\zeta) = \begin{bmatrix} \left[R' Q_{ZZ} R\right]^{-1} R' Q_{ZX} \\ Q_{XX} - [Q_{XZ} R (R' Q_{ZZ} R)^{-1} R' Q_{ZX}] \end{bmatrix}$. To calculate the derivative of $b(\zeta)$ with respect to ζ we use the differential of $b(\zeta)$:

$$\mathrm{d}\{\left[R' Q_{ZZ} R\right]^{-1} R' Q_{ZX}\} = \mathrm{d}\{\left[R' Q_{ZZ} R\right]^{-1}\} R' Q_{ZX} + \left[R' Q_{ZZ} R\right]^{-1} R' \mathrm{d}\{Q_{ZX}\} \qquad \text{(A.12)}$$

The differential of first term in the RHS of (A.12)

$$\mathrm{d}\{\left[R' Q_{ZZ} R\right]^{-1}\} R' Q_{ZX} = -(R' Q_{ZZ} R)^{-1} R' \mathrm{d}\{Q_{ZZ}\} R (R' Q_{ZZ} R)^{-1} R' Q_{ZX},$$

taking the vec of both sides

$$\mathrm{vec}\Big[\mathrm{d}\{\left[R' Q_{ZZ} R\right]^{-1}\} R' Q_{ZX}\Big] = -\Big[Q_{XZ} R (R' Q_{ZZ} R)^{-1} R' \otimes (R' Q_{ZZ} R)^{-1} R'\Big] \frac{\partial \mathrm{vec} Q_{ZZ}}{\partial \zeta'} \mathrm{d}\zeta.$$

Thus, denoting $Q_R = (R' Q_{ZZ} R)^{-1}$,

$$\mathrm{d}\{Q_R R' Q_{ZX}\} = \left\{-\Big[Q_{XZ} R Q_R R' \otimes Q_R R'\Big] \frac{\partial \mathrm{vec} Q_{ZZ}}{\partial \zeta'} + Q_R R' \frac{\partial Q_{ZX}}{\partial \zeta'}\right\} \mathrm{d}\zeta.$$

Now, the differential of the second block of rows in $b(\zeta)$

$$\mathrm{d}\{Q_{XX} - [Q_{XZ} R Q_R R' Q_{ZX}]\} = \mathrm{d}\{Q_{XX}\} - \mathrm{d}\{[Q_{XZ} R Q_R R' Q_{ZX}]\}$$

with $\mathrm{d}\{Q_{XX}\} = \frac{\partial Q_{XX}}{\partial \zeta'}\mathrm{d}\zeta$, and

$$\mathrm{d}\{[Q_{XZ}RQ_RR'Q_{ZX}]\} = 2[Q_{XZ}RQ_RR'\mathrm{d}\{Q_{ZX}\}] - Q_{XZ}RQ_RR'\mathrm{d}\{Q_{ZZ}\}RQ_RR'Q_{ZX}.$$

The second term on the RHS of the previous expression can be rewritten, using the trace operator, as

$$Q_{XZ}RQ_RR'\mathrm{d}\{Q_{ZZ}\}RQ_RR'Q_{ZX} = \mathrm{vec}(Q_{ZX}Q_{XZ})'[(RQ_RR')\otimes(RQ_RR')]\mathrm{dvec}(Q_{ZZ}).$$

The Jacobian matrix of $b(\zeta)$

$$\frac{\partial b(\zeta)}{\partial \zeta'} = \begin{bmatrix} -\Big[Q_{XZ}RQ_RR'\otimes Q_RR'\Big]\frac{\partial \mathrm{vec}Q_{ZZ}}{\partial\zeta'} + Q_RR'\frac{\partial Q_{ZX}}{\partial\zeta'} \\ \frac{\partial Q_{XX}}{\partial\zeta'} + \mathrm{vec}(Q_{ZX}Q_{XZ})'[(RQ_RR')\otimes(RQ_RR')]\frac{\partial \mathrm{vec}Q_{ZZ}}{\partial\zeta'} - 2[Q_{XZ}RQ_RR']\frac{\partial Q_{ZX}}{\partial\zeta'} \end{bmatrix}.$$

In order to prove that the Jacobian matrix has full-column rank, i.e. equal to 4, we focus on the matrix (as in the Proof of Proposition 4) $C = \begin{bmatrix} \frac{\partial \mathrm{vec}Q_{ZZ}}{\partial\zeta'} \\ \frac{\partial Q_{ZX}}{\partial\zeta'} \end{bmatrix}$. This matrix contains the partial derivatives of the variance and the autocovariances of the RV_t in the case of the Heston model. Let denote c_j the row of C which contains the partial derivative of $\gamma(j)$, i.e. $c_j = \begin{bmatrix} \frac{\partial\gamma(j)}{\partial\kappa} & 2\mu + \frac{\partial\gamma(j)}{\partial\omega} & \frac{\partial\gamma(j)}{\partial\varsigma} & 2\mu\Delta + \frac{\partial\gamma(j)}{\partial\sigma_u^2} \end{bmatrix}'$. The matrix C has reduced column rank if there exists a vector $w = [w_1, w_2, w_3, w_4]' \neq 0$ such that $c_j'w = 0$, and this must hold for all rows of C, that is $Cw = 0$. Since $\frac{\partial\gamma(j)}{\partial\sigma_u^2} = 0 \quad$ for $j > 0$, we have that for $j > 0$

$$w_1\frac{\partial\gamma(j)}{\partial\kappa} + w_2\left(2\mu + \frac{\partial\gamma(j)}{\partial\omega}\right) + w_3\left(2\mu\Delta + \frac{\partial\gamma(j)}{\partial\varsigma}\right) = 0 \tag{A.13}$$

and for $j = 0$

$$w_1\frac{\partial\gamma(0)}{\partial\kappa} + w_2\left(2\mu + \frac{\partial\gamma(0)}{\partial\omega}\right) + w_3\frac{\partial\gamma(0)}{\partial\varsigma} + w_4\left(2\mu\Delta + \frac{\partial\gamma(0)}{\partial\varsigma}\right) = 0. \tag{A.14}$$

Now equating (A.13) and (A.14), we get

$$w_1\Big[\frac{\partial\gamma(j)}{\partial\kappa} - \frac{\partial\gamma(0)}{\partial\kappa}\Big] + w_2\Big[\frac{\partial\gamma(j)}{\partial\omega} - \frac{\partial\gamma(0)}{\partial\omega}\Big] + w_3\Big[\frac{\partial\gamma(j)}{\partial\varsigma} - \frac{\partial\gamma(0)}{\partial\varsigma}\Big] - w_4\frac{\partial\gamma(0)}{\partial\sigma_u^2} = 0, \tag{A.15}$$

where the expressions in parenthesis are functions of j. This means that the value of w_4 which satisfies (A.15) is

$$w_4 = \Big[\frac{\partial\gamma(0)}{\partial\sigma_u^2}\Big]^{-1}\left\{w_1\Big[\frac{\partial\gamma(j)}{\partial\kappa} - \frac{\partial\gamma(0)}{\partial\kappa}\Big] + w_2\Big[\frac{\partial\gamma(j)}{\partial\omega} - \frac{\partial\gamma(0)}{\partial\omega}\Big] + w_3\Big[\frac{\partial\gamma(j)}{\partial\varsigma} - \frac{\partial\gamma(0)}{\partial\varsigma}\Big]\right\},$$

which is obviously not constant since it depends on j. Thus the only vector w which satisfies (A.15) is the null vector. We can conclude that the matrix C has full column rank.